# Participant 1 (P1)

Interviewer: So, thank you for coming.

Participant: No problem.

Interviewer: How would you describe your experience level with computer vision and NLP?

Participant: Probably not expert, but I have read a handful of papers from both and a handful of papers involving neural network diagrams. More than a handful.

Interviewer: What is your role at the moment?

Participant: PhD student. First year.

Interviewer: Can you describe how you use diagrams when communicating your research?

Participant: Ok, so I have yet to publish any research, but in terms of communicating with colleagues about ideas about research, I often would use the diagram as a schematic overview of a process, or if I had a particular idea for an architecture, yeah just to convey the kind of mathematical function that is represented by that architecture.

Interviewer: How do you decide what level of granularity to draw your diagram?

Participant: It definitely depends on the context, so it depends on who I’m speaking to, and at what stage the idea is at. So if I just have a general idea of what kind of technique I want to use to solve a problem, then it wouldn’t be very detailed, but if it is something I know I want to implement or have already implemented then it would be more detailed.

Interviewer: What tools do you use for making diagrams?

Participant: So, I use drawing programs like google draw, but I’ve also used tix, the LaTeX diagram package as well.

Interviewer: And how do you use diagrams when consuming research?

Participant: [Pause] A few different ways, like the two aspects that I was saying before, wide schematic overviews of processes, to give a visual representation for a block of text that just explained a process, but then also in more precise mathematical sense, particularly for neural network diagrams, so I can better understand what the function is, what the function is doing, what the network is doing.

Interviewer: So is this about reading a paper?

Participant: Yes

Interviewer: So is there a problem in consuming research more broadly? So are you able, for example, to read what you want from the diagrams?

Participant: In most cases yes, I mean, usually not solely from the diagrams. It is a process of moving to and from the text and the diagram to get a complete picture. Both of them at once is how I get the best understanding.

Interviewer: Do you print them out?

Participant: Particularly for the diagrams or just do I print papers out in general?

Interviewer: I wondering about [Pause] so you described an interaction between the diagram and the text, and I’m wondering about how you are flicking through?

Participant: I wouldn’t say I’m consistent, but yeah it is either flicking from page to page or just scrolling.

Interviewer: Thank you. The next part is a series of diagram examples. There are six diagrams here, they are in a bit of an order. Tell me about it, and what you like and don’t like about it.

Participant: [A Pause] Ok. There are a lot of things I like about this one. First of all, this representation of the dimensionality of different things, you immediately know this embedding layer is a matrix and these things are vectors, and get a heuristic sense of how big they are, and I like the notation of arrows pointing to mathematical operators, gives you a clear idea of what is going on. Yeah and this labelling along the axis is helpful and gives an overview of what each of the components is supposed to be doing as part of the wider network. I guess that it is colourful is nice, nicer to look at than a black and white diagram.

Interviewer: Anything you don’t like about it?

Participant: [Pause] I’d say, I mean it looks a bit crowded in terms of how many arrows there are, but I think that is kind of necessary, I’m not sure there is any way around that. No, I’d say it is a good diagram, if you ask me.

Interviewer: So next one? B.

Participant: Fig 1?

Interviewer: The captions are from the papers.

Participant: Ah ok. [B Pause] Ok so, one thing I don’t immediately understand about this diagram, probably because I don’t have the context of the wider paper, I don’t know what a Siamese structure is. I guess I can take a guess from what it says here but I don’t exactly know what that means from looking at the diagram. [Pause] And I don’t really understand what is the difference in the inputs from the two streams. The inputs are labelled the same, so it is not clear what the difference is, or why there are two streams in the first place. This middle block that is coloured dark blue with white silhouettes, obviously has some particular role. I can get that from the caption, that it is the person segmentation going on, but something over the top, like this is where the person segmentation happens, would potentially be helpful. [Pause] Other things, these arrows here are kind of a bit narrow. Obviously the fact it is bidirectional arrow is important but the arrowheads are extremely small. It could be perhaps more detailed about what contrastive loss is. Mathematical information might be helpful but I guess that is in the rest of the paper. It does give a clear impression of the dimensionality of everything and how that is progressing as it goes through the network. So it says “all the convolutional layers are shared” but kind of these first ones in yellow and green, there’s a arrow later saying they are shared, but the caption implies all of these are shared, and I would kind of expect an arrow in that first region to indicate that, if that is the case.

Interviewer: Thank you. So the next one, C.

Participant: [C Pause] In this diagram is not immediately clear what the inputs and outputs are from a first glance at it, because you’ve got the stereo input images and the point clouds going in. Maybe they are both inputs, and the outputs, I guess it is this thing in the bottom right, but maybe it would be more helpful if the progression in the network went left to right, as at a first glance it isn’t clear which way to navigate round. There are some links which aren’t explained. There is quite a lot of notation, mathematical notation, which is in the diagram but not explained in the caption. I don’t know what half, any, of these symbols stand for. [Pause] I don’t really know what the colouring of the arrows represents, if anything, or if that is just an aesthetic choice. I think the main thing I would say that is not good with this one is it an enormously complicated network by the looks of things, but not very detailed in the caption, this long tailed distribution is not even mentioned in the caption I believe. A lot of components are not explained in terms of their role in the network. Or what the iteration process is.

Interviewer: What do you like about it?

Participant: [Pause] I like that the input is actually an example, not just a representation of a matrix that represents a picture, so you can actually see an example of what one of the inputs would be, and the same with the outputs, that’s something that is nice.

Interviewer: Thank you. D now.

Participant: [D Pause] So… I don’t really understand some of the arrows in this kind of first bit, the NL encoder. Presumably there is the embedding [Pause] taken as an input by the encoders. And then, there are bidirectional arrows on the encoder, but [Pause] I’m not sure what that means in this context. [Pause] Similar to the last one, it is clear to me what the input is, natural language is to be interpreted as an SQL query, but they’ve got numerous outputs but to me the caption seems to suggest it is just trying to synthesise a SQL query, but they’re kind of got this memory table in the top right. [Pause] Overall, there’s a lot going on, and again, not very much explanation. [Pause] And yeah, a few mathematical bits of notation are not defined, like the H subscript X, B subscript T and stuff like that. Yeah, overall pretty confusing.

Interviewer: Anything that you like about it?

Participant: Yeah again I like that it is displayed with an example, but I guess by necessity, the SQL query in natural language will be a long sentence, but it is confusing having that many words in a sentence that is input, it makes for a large diagram. I think if it was accompanied by a more succinct diagram, just an overview of the different parts, it might be clearer if it had that as well as this one with an example included. Yeah, other than that, not much I like.

Interviewer: E?

Participant: [E Pause] There are a couple of good things and a couple of bad things about this one. The main bad thing being the caption, which may as well not be there, because it doesn’t tell you what the model is doing, or tell you the role of each of the constituent parts they’ve divided it up into, which would be nice. But it is quite clearly segmented in the diagram into each of the roles. Just having the titles of the roles, and a bit of explanation in the caption would be nice. Again, I like it has the pictures of the example inputs. [Pause] I’m not really sure about the use of colour and what this is supposed to represent, for example, the two kind of sections labelled in green with green outline, are they supposed to convey similarity between those two or is it kind of random, the way they’ve coloured them? But they are quite clear, the little subdiagrams. And it’s kind of clear what the inputs and outputs are from the diagram.

Interviewer: So for this one, for example here we have 10 nodes or features or something there. Do you think it reflects precision in terms of elements?

Participant: I doubt it, I wouldn’t expect so. [Pause] It would be good to have a labelling of the dimensionality of the different layers, the same with the embedding layer aswell. [Pause] It’s a bit unclear here, there are three different outputs here, in transcript embedding, but only one is propagated to the context aware fusion module. [Pause] And again the link between context aware fusion module and the procedure extraction module we go from something that looks like a vector to something that looks like a feature matrix. It could do with a labelling of what mathematical operation are happening, is it concatenation of different features into a feature matrix, that would be helpful. [Pause] It looks like it could be a good diagram if it had a lot more explanation in the caption of what was happening between the subcomponents of the network.

Interviewer: Cool. And the final one?

Participant: [F Pause] So I kind of like this one, that this one is very concise, small diagram of something that has quite a few components that summarises what is going on quite well. But again, very undetailed caption, like the merging layer, I’m not sure what the merging operation is from just the diagram and caption, presumably that is explained elsewhere, it would be good to have it so the caption kind of stands alone. [Pause] Kind of unsure what the shapes and outlines of shapes are supposed to be like. Why is merged in a different kind of box to everything else. There is no explanation of how merging is different to anything else, in a rectangular box. You can’t really tell what the network is doing, there is very little context. I just know it takes a summary and text and produces a score but what does the score represent, it just lacks the wider context, but concise and yeah, but yeah, doesn’t give you a good idea of what’s going on mathematically, what the dimensions are, as a result of being so concise, you couldn’t have all that going on and it being so concise. Again, another diagram with an example, that kind of showed dimensions of the stuff might be helpful. The lack of colour [Pause] doesn’t really detract from understanding of the diagram. Yeah. Yeah, I’m not sure, I don’t think it would benefit from colour.

Interviewer: Are you hinting towards aesthetics?

Participant: I’m just thinking about why people use colour in diagrams. I guess this is so high level you don’t really need to use colour to differentiate different aspects, as it is very much a summary of the network.

Interviewer: Have you seen any diagrams like these before?

Participant: Yeah definitely, usually on the high detail side than this final one, but I think these are helpful alongside a more detailed diagram aswell. Definitely see lots of these enormous diagrams with lots going on and examples.

Interviewer: You’re pointing to E and A

Participant: Yes, particularly the fifth example typifies it.

Interviewer: You described your priorities in terms of details you like, inputs and outputs. Is there anything else you are trying to get from the diagrams?

Participant: One thing I don’t like about looking at diagrams is if there is an overload of stuff, in terms of arrows and labels for things. But in a lot of cases that is necessary to see what is going on. In terms of this final diagram, it is a lot easier if you at first glance, less intimidating [sic] a diagram than this one, it is going to take me a while to work out what is going on. I’ve got a good idea of the constituents and what the inputs and outputs are like straight away. Definitely quicker to process which is good.

Interviewer: Thank you. Umm. Just one last question on this. What do you think about the variety of diagrams used in neural network style papers?

Participant: I would like more standardisation. Just because every time you get a new diagram you have to learn that particular authors way of trying to communicate via a diagram, every time. And I guess there are some standards that are, you know, kind of used widely but it is like you’re trying to learn [Pause] how to understand, I don’t want to say a new diagrammatic language, but a new way of communicating every time you see a new diagram, a lot of the time.

Interviewer: Anything else you’d like to say about these examples before we move on?

Participant: Umm, no I don’t think so.

Interviewer: Cool. The next part is a card sorting exercise. The question is to rank these in order of how important they are in your use of research paper diagrams. You can add or discard any that you wish. The top three or four or so are the most important.

Participant: [Pause] Sorry this is not me ranking them this is just me thinking. So, when you say index to navigate the paper, does that mean like an anchor, was it before or after the diagram this bit of text?

Interviewer: It means whatever you want it to mean, my intention was more like a table of contents to say “oh this is the overall architecture, I can go look up that bit”.

Participant: [Pause] Do you want me to think out loud while I’m doing this?

Interviewer: That would be excellent.

Participant: Does the corpora and data types include inputs and outputs?

Interviewer: Yeah, but if inputs and outputs is important let’s make an extra card for that?

Participant: Yeah, yeah.

Interviewer: These cards are just to frame your thinking, not rigid.

Participant: [Pause] It is very much contextual, is this like if I’m seeing a diagram for the first time, or if it’s papers I will refer back to often, and I know the inputs and outputs and the purpose, but I want the diagrams for another purpose. My requirements change based on how familiar I am with the particular system, by system I mean network.

Interviewer: I think it would be interesting to explore both, the highly familiar and highly unfamiliar?

Participant: For a paper that I am reading for the first time, a paper that might be useful for my research, inputs and outputs, definitely. Definitely important, purpose of the system. I don’t really think about the author’s intention or altering it when I look for the first time. External dependencies. Is that like software dependencies? Like just anything else.

Interviewer: Whatever you like. I think I had in mind like wikigold or other external resources.

Participant: Yeah so system novelty, contribution. Understanding how works is definitely top, and I do that by inputs and outputs and specific architectural features, identifying the layers that tells me how it is working. If I can put this one together with these, that’s what I’m trying to do. I’m trying to understand how the system works by identifying inputs and outputs, identifying layers, relations between components, and specific architectural features. That’s one thing I’m trying to do when looking at a diagram. Initial check to see if I’m interested, if it is architecturally interesting. I don’t really think about it a memory aid, nor as an index to navigate the paper. The diagram usually corresponds to a section of the paper. Yeah, comparing to other systems, depending on if I’ve seen other networks for that fulfil the same purpose as this particular one. Representational choices I’d throw in with understanding how the system works. Corpora and data types, [Pause] that’s maybe something you might want to find out from a diagram. I don’t really think about external dependencies. I wouldn’t say an aid for writing a summary, maybe I would, but more if I revisit a paper, having read it once, I’ll just look at the diagram. So that’s it for a first look at a paper.

Interviewer: Could you just read through to summarise?

Participant: Yeah sure, top I lumped a number of things under a “understanding how the system works” umbrella, and to do that, identify the inputs and outputs, identify the layers, relations, identify representational choice, specific architectural features, then after that, separately, identifying the purpose of the system, then probably an initial check to see if they use a particular thing I am interested in, that’s very situational, umm, depends on whether I am interested just for the architectures sake or for the task. Yeah systems novelty, comparing to other systems, memory aid for what the papers about. Corpora and data types, I guess the memory aid one is what I was trying to say there. [Pause] Then the less important ones, I don’t really think about the authors intention so much when I look at a diagram, and when I look at it for the first time I don’t look at altering the architecture or as an index for the paper. In general most of the diagrams that I look at don’t have external dependencies. If I was looking at a paper that I knew very well, that was very important for my research, so not just trying to get a gist of the paper but trying to properly understand it and use the architecture or modify it in some way then I’d move up altering the architecture. And in this context all this how the system works stuff it is very important the amount of detail that is there. If I’m trying to implement the diagram in code I need to know the dimensions of all the things which constitute the network. [Pause] If it was a diagram with external dependencies I would be looking at the diagram to find out what external dependencies there were. Those are the main differences

Interviewer: Fantastic, thank you very much. The final part is just some catchall questions. Do you use diagrams to reason as part of your creative process when doing research?

Participant: [Pause] Yes, I do in kind of the sense of if I’m trying to solve a task in some kind of way to depict that visually I would separate it out into stages of the task. Just a sketch, like a box, very much like the final diagram, but more abstract, representing phases of the process rather than specific layers.

Interviewer: Your gestures… is that pen and paper?

Participant: Yeah, yeah.

Interviewer: And it’s before or while coding, or?

Participant: Oh definitely before. Usually it’s in the context of I’m thinking of a new way to solve a problem, I’ve run a bit of maths down or something, and I’m trying to bring it all together in my head, what the solution I’m actually thing of is, it is kind of hard to look at a page of maths and look at the broad process, a broad description of the process, you get too into the details, this helps to summarise in a way.

Interviewer: So you’re using it as a abstraction tool to group things together?

Participant: Yeah and to make it clear to myself what I’m trying to do here. Group together, put into components.

Interviewer: You’re calling this a diagram, not a list, so does it have some spatial arrangement?

Participant: I’d say usually it is like a directed feed forward, this then this then this.

Interviewer: Cool, thank you. Do you use diagrams in any other way in your research more broadly?

Participant: [Pause] More broadly than neural networks?

Interviewer: More broadly than the creative process.

Participant: As in diagrams that I produce? A lot of literature I read contains diagrams that are not neural network pictures, like an overview of processes showing what’s going on, I feel they are quite useful in textbooks and stuff like that. Now I’m back in particle physics, Feynmann diagrams are back in my life and are a whole topic in themselves I guess.

Interviewer: Can you think of any tools which would help you in making or consuming diagrams?

Participant: [Pause] When a diagram is too detailed in a paper, you’re constrained by the fact it will be printed in a pdf, it can’t be interactive, you can’t hover a mouse over an have an explanation. There are diagrams like that on the internet that help with the information overload that you get when you kind of look at a diagram in a paper. Obviously in scientific publications you are constrained not to do that. That’s one thing, preserving space, or being able to vary the amount of detail you get about a particular section of a diagram might be helpful. Diagrams in non-pdf format would be good for consuming diagrams. That’s about consuming diagrams. Creating diagrams, [Pause] a style guide might be helpful, or something like that. I don’t just mean style, some more you know conventions and generally agreed conventions on how to depict different components, might make things easier both on the creation and understanding, cos you do have to think a lot about how you’re going to represent this thing.

Interviewer: Have you used tensorflow?

Participant: Essentially no, I pretty much exclusively use pytorch. I did use keras very briefly a couple of years ago, but no, pytorch.

Interviewer: Is there anything else you’d like to say about diagrams while you’re here?

Participant: No, not really, I like diagrams though.

Interviewer: Brilliant, thank you very much.

Participant: No worries.

# Participant 2 (P2)

[Being familiar with the participant, the interview started with some informal conversation which has been omitted]

Interviewer: Thanks for coming to this interview. If you could just describe what your role is at the moment, and what is your experience with natural language processing and computer vision?

Participant: I’m a PhD student in natural language processing, did a masters in cognitive science and IT, where I was doing similar things, I did some vision courses there but definitely not an area I am any kind of expert in.

Interviewer: You’re in your final year?

Participant: Yeah in my final year of my PhD.

Interviewer: And that is primarily on natural language processing?

Participant: Yeah it is more applications based, but yes.

Interviewer: Excellent. Can you describe how you use diagrams when communicating your research?

Participant: [Pause] Not that much, but the diagrams I have used are usually of pipelines, usually of natural language processing pipelines that I’ve drawn myself. A few papers I’ve had to draw a diagram to explain some kind of pipeline of the process or processing.

Interviewer: Why do you use a diagram to do this rather than an algorithm or text?

Participant: Probably got quite a lot to do with my background, which is less technical, I’m quite a visual person, I’ve got a degree in art, [Pause] so I personally find that a lot easier to understand myself.

Interviewer: Cool. What sort of problems do you encounter when creating these diagrams?

Participant: [Pause] I don’t know, can’t think of any particular problems. It is just a process of trying to simplify something in a way I think people can understand easily. But I don’t find that particularly problematic, usually.

Interviewer: So how do you decide what to include in your diagram?

Participant: [Pause] I don’t know, I don’t have a systematic method. I guess I’ve always got an idea about what a figure is trying to communicate that I feel is easier done in images than words but it just depends what that happens to be. I don’t follow any kind of systematic method.

Interviewer: And does that also hold for how you represent it?

Participant: [Pause] Yeah I would say so, I guess I am inspired by other diagrams that seem to communicate well to me. I don’t have a specific [Pause] diagram maker to point to that I think does a great job.

Interviewer: Do you have a tool that you tend to use to create diagrams?

Participant: Yeah I usually use Illustrator for most things, been using draw.io recently, but I find it a little bit clunky.

Interviewer: How do you use diagrams when consuming research?

Participant: [Pause] As I say I find visual representations a lot easier, usually, than either pseudocode or long prose descriptions, so I’m usually pretty happy if a diagram has been included of anything that might be a bit complex to understand. Or even, I guess a lot of the time I am looking at that before I look at the text, if there is a diagram present.

Interviewer: Interesting, thank you. I think we can probably move on to this diagram series now. This is a series of six different diagrams. If you could please tell me about this diagram and what you like and do not like about it.

Participant: [A Pause] Ok, so it’s obviously some kind of neural network architecture. I would say it is fairly clear, though I’d say I’d definitely have to have seen a lot of things like that before you could interpret it. If I was completely new to this then it would not be great, there is maybe information missing, this might require text as well I guess. It is not clear the actual what it is for, it’s there but there is nothing apart from the title I guess. Yeah, it’s pretty clear, and the use of colours and stuff helps to separate things out. There are some arrows and stuff that probably it’s not clear unless you have some text explanations of. Anything else required for that?

Interviewer: You mentioned you might have seen things that are like this before?

Participant: Yeah

Interviewer: Is this quite a common pattern?

Participant: Yeah, it follows the layers, you can see clearly there are different layers, and a sigmoid function at the end. I guess it flows the way most of these things are drawn. Though there is no systematic way of going left to right, up or down, or top to bottom or anything like that, but it follows the general pattern.

Interviewer: Do you have a view on the mathematical notation that is used?

Participant: [Pause] Yeah, I would need some supporting explanations of some of this. [Pause] I think it is fairly standard.

Interviewer: Cool, thank you. Next one?

Participant: [B Pause] Yeah. Ok this is less familiar, the way this has been drawn. [Pause] So, just read the title. Yeah so I’m not so familiar with vision stuff, I guess that’s why the 3D stuff is being used, they need 3D diagrams. Yeah I guess I find this a little harder to interpret, it is something I’m less familiar with. I also, I’m not too familiar with Siamese networks, though I recognise the term.

Interviewer: So these numbers on these?

Participant: Yeah so that’s something to do with the [Pause] windows of like pixels or something? You know I haven’t done deep learning for vision, so I would need probably to read the paper or have a bit more background to interpret that better. I find the graphics pretty ugly, but that’s my design background.

Participant: [C Pause] Yeah there’s a lot of things I don’t know what’s going on here, it’s outside my field. So, little complicated to work out, there’s things going on in different directions which I don’t know if they’re simultaneous, I’m finding it hard to interpret. Again it’s kind of a vision thing. [Pause] Yeah it is not really clear to me what the outcome is, I’d definitely need to read more on that to be able to interpret the diagram.

Interviewer: I’m not going to go into the technical parts, just the visualisation mechanism.

Participant: Oh ok.

Interviewer: This is using an example throughout, I’m interested in whether you think that’s helpful?

Participant: I can see the example on this side over here on the right, but I have no idea about what’s going on with these clouds, I have no idea about the terminology or the notation, or what those images actually are, I can’t relate them to the input image particularly.

Interviewer: Cool, thank you, that’s the end of the computer vision ones, so.

Participant: [D Pause] Ok, so quite a lot going on in this one as well. Initially quite hard to interpret cos input is at the bottom, and naturally you look at stuff at the top first. [Pause] I haven’t seen this kind of problem either. I don’t really know what to say about that one. Have you got a specific question?

Interviewer: Is there anything you like about it?

Participant: [Pause] Seems a little bit messy to me, I don’t really know because also this is not a problem I work on I don’t really know where to focus my attention on this. It is an overview of the model, but as I don’t have a specific question about that model, it explains what’s there but it was hard to interpret for me.

Interviewer: Again here, we have what looks like an example being used throughout?

Participant: Yeah, againit took me a while to know where to look to find something. I’m looking for some input and output and it seems to be the opposite way round to my natural way of interpreting the diagram.

Interviewer: Great, thank you.

Participant: [E Pause] Yeah this is automatically easier, because you’ve got the input on the top left, and it’s a bit clearer about where you’re supposed to follow the model, as you’ve these clear arrows and guidelines. [Pause] I quite like the design on this one, it is relatively easy to follow due to the structure of it.

Interviewer: You were indicating the modules?

Participant: Yeah, each part is very clearly separated and you’re told how to follow it by the arrows. Without having worked on this problem you can see what’s going on quite easily.

Interviewer: In the context aware fusion module, we have 10 things here. Do you think that is significant, the 10?

Participant: Well, it’s not clear whether that is just an abstract representation, I’d have to look in the text, it is quite an abstracted image, unless it says in the title or something that it was significant I wouldn’t interpret it as being significant, or I’d look in the text. Whereas in the embeddings, it specifically says there are an unknown number of layers.

Interviewer: [Pause] The example in this one, is that helpful or is it not necessary because the flow is clear?

Participant: I think that’s helpful, it puts it into context of what the task is. [Pause] Yeah that’s relatively clear I think.

Interviewer: Is there anything you don’t like about it?

Participant: Well, so, I mean, so you’ve got that example which is quite clear, but then you have to look over to the entire diagonal opposite to understand what the task is. I mean maybe you could guess it, but it could say it earlier. Overall I think that is a pretty clear diagram compared to some of the others. Just more nicely designed, easier on the eye, and easier to follow.

Interviewer: When consuming diagrams, is aesthetics very important?

Participant: Yeah, in terms of clarity, [Pause] I guess sometimes you’re looking at them for different reasons, whether you’re skimming or reading a paper just to get the idea of a paper, and sometimes you’re actually looking at the details, so that might effect what I’m looking for in the diagram depending on what I’m looking for at that particular time. In general the clarity is more important than the details just to give me a general idea of what’s going on.

Interviewer: Ok, thank you.

Participant: [F Pause] This is fairly simple. I can follow most of it pretty easily, although it maybe want a more [Pause] explicit explanation of the task, because I’m having to guess that from what the input is, and then the output just says score so it’s not entirely clear, there’s some kind of summarisation or something. The bit in the middle, without reading the text I’m not sure what’s going on there. There’s a lot of information not here unless you’ve read the text. [Pause] But the simplicity makes it easy to get the general idea very quickly. I think most of these haven’t had enough information in the caption, certainly without having read any of the text so I don’t know why it has been merged, what the significance of mean over time is or exactly what the task is, I need further information.

Interviewer: [Pause] This labelling they have for the layers, does this make sense to you or match with your artistic preferences?

Participant: Yeah, it is fairly straightforward and I think that makes perfect sense.

Interviewer: Anything else to say now we’ve had a look at these diagrams? You made some comment about the captions?

Participant: Captions could be fuller, there’s a reason captions are given with diagrams and figures so they should be made use of. Often when skimming a paper you are just looking at the kind of maybe[sic] just read the abstract, then just flick through the paper and look at the pictures. So if there is something to explain that without having to find where in the text, that would be good. Some of these are pretty complex, without knowing them or reading the text, but then it’s hard to know. Maybe they need to be complex, without having seen the text.

Interviewer: Thank you. The next task is about tasks, it is a card sorting exercise, if you could please rank these in order of how important they are in your use of research paper diagrams. And if there’s something that’s missing or [interruption]

Participant: So diagrams [Pause] How important it would be to have a diagram for this?

Interviewer: So you’re consuming some diagrams in research papers, what tasks are you performing? What are you trying to get from the diagram, or more generally how important are they in your use of it, it may not just be information extraction.

Participant: Right ok. Not sure how easy this will be, so many options!

Interviewer: What would be helpful is if you get the top few right, and then it would perhaps descend. If you’re able to explain your thinking, or talk aloud as your doing this, that would be great.

Participant: Ok. Talk as I do it, yeah ok. Not sure if this is going to be, how final this is. Index to navigate a paper, I guess this is the task that we’ve been talking about the most, it’s just trying to find out quickly what the papers about, is it relevant to me, that’s pretty common I’d say. I’ve also put identifying specific architectural features, so this would be more like if I’m trying to reimplement or do something similar to something I’ve seen in a paper, as I’ve said before I often find it easier to look at pictures than prose or pseudocode to find out what I actually have to do to replicate something. [Pause] Initial check to see if they use a thing I’m interested in, I guess that is a subtask of navigating the paper in a way. Just a general impression of the paper. [Pause] Understanding how the system works, that also goes with identifying specific architectural features. [Pause] Yeah not too sure about most of the others. I don’t often find myself writing summaries, well maybe if you include literature or background work it’s maybe, but I don’t think at this stage it’s normally too much focus on the diagrams. [Pause] There’s a couple here that for me are kind of similar. Identifying the specific architectural features, understanding how the system works, identifying layers and relations between. Just understanding the architecture of the system, for me those go together. Again, representational choices might be something I do sometimes. [Pause] Yeah. I’d say these two groups.

Interviewer: Could you just read those out?

Participant: Navigating the paper and checking to see if it’s relevant, I guess they can be interpreted as different but I’ve interpreted them as similar. And then just identifying parts of whatever is being implemented in the thing for the purposes of trying to build on the thing or trying to do something similar. Do you want me to go any further with these?

Interviewer: If there’s anything you’re like gosh I don’t do that, then that would also be interesting.

Participant: I don’t know about this memory aid. I think, not sure, I don’t think I use diagrams as a memory aid. I guess identifying what the author thinks is important might again just go with navigating the paper and that initial reading of the paper, because you do assume that what they’ve taken the bother to make a diagram of is going to be important. So I’d stick that in with those, or near them anyway. And there there’s a few identifying ones that I guess I do too. Yeah.

Interviewer: Thank you very much. So the last section is just some catchall questions. Do you use diagrams to reason as part of your creative process when doing research, and if so how?

Participant: [Pause] That’s a good question. Yeah, yeah I do sometimes, probably more to explain an idea that I’ve already had to someone else, rather than to reason it out to myself. [Pause] Yeah I would say more of that. Although I definitely make sketches occasionally of ideas.

Interviewer: Pen and paper?

Participant: Yeah pen and paper. Like in a notebook of some pipeline or whatever might work. And then yeah definitely to explain.

Interviewer: Would you do that also on pen and paper?

Participant: Yeah usually, although I might then if it’s going to go a bit further I might some sort of like vector graphic or something, if it’s going to be something a little more formal, if I’m going to send it to my supervisor or something. Although yeah, pen and paper is more common.

Interviewer: Do you use diagrams in any other way during research more broadly?

Participant: Any other way than creative process, writing papers [Pause] and explaining things, I guess that’s more or less it. Presenting work. Posters and slideshows and things. I make sure of diagrams there.

Interviewer: Final one, can you think of any tools that would help you in making or consuming diagrams?

Participant: [Pause] I don’t know I’m pretty happy with the tools that I use really. It may help if there is more of a, not really a tool, but a more systematic way of doing things that certainly in terms of consuming diagrams. It’s up to each researcher. I guess I like the creativity of that, and sometimes I feel like sometimes it is done quite badly and are difficult to interpret, and people might benefit from a more systematic way of creating diagrams. I’m not sure, I haven’t seen it work. Then, tools, you mean like just some kind of diagram drawing software? I’m happy with what I use.

Interviewer: What you were talking about there in that you’re consuming diagrams and that there’s different types that you have to get your head around [Interruption]

Participant: I think that people could benefit from some design knowledge about the way that people interpret things, for eye tracking that generally it is better to start in the top left if you can, and to make things flow in a way so you don’t really have to think about it. And obviously most people haven’t been trained in that, I’d imagine.

Interviewer: Since you have an artistic background I’m interested in digging a bit more, into the use of language, the use of text within the diagram. Do you have thoughts on this?

Participant: In general, I like to see things simplified as much as possible and not have anything that can be omitted but still make sense should be omitted. I guess it’s what’s best is when the diagram and the text work well together. So, you should be able to get the general idea and refer to some ideally clearly written text which is quite close on the page to the diagram, and those two should work quite well together so you don’t have every detail in the diagram. I guess it is quite nice where you can leave text out but then obviously text should be used to highlight important things. Yeah.

Interviewer: So the use of colour for example, do you think that is part of this graphical design thing that people need to improve?

Participant: You need to consider accessibility these days, colour certainly helps if you have access to colour, but not everyone will, or want to people want to print out stuff out and only have black and white, or might have some colour issues you might need to consider. Colour can help but you always need to consider your audience when you use colour, so.

Interviewer: I’m just going to go slightly off-piste and go back to some of these examples if I may, so the fifth example, this one, so you spoke just then about omitting as much as you can. I’m thinking about how schematic this is, these layers 1 and n, they’ve omitted layers 2 to n-1, whereas in this iconic representation just above that, they’ve not made that representational choice, and throughout the diagram they’ve made different representational choices. I’m just interested in what your thoughts are on that.

Participant: Well, I don’t know yeah without knowing I might assume there is something important about the things they have clarified on the diagram that are particular to what they are doing.If the author is going to have that kind of inconsistency you’d hope that there is a good reason for that, and they think I don’t need to know the details of the part that is omitted.

Interviewer: In this same example, we have an ellipsis being used to indicate missing information, and this is not done here for example

Participant: Right yeah

Interviewer: Do you feel like there is a problem in inconsistency in that kind of encoding?

Participant: No I don’t think so, because they’re different things. The text example is more or less just a kind of visual representation of what the input is. That inconsistency doesn’t bother me because it is across two different things. If there was another module with more neural network layers where they were using those ellipsis dots then that would be jarring, but I don’t think it does.

Interviewer: For the first example, they use this ellipsis and this k subscript notation

Participant: Yeah

Interviewer: Do you feel it benefits from having both of these things? Would you prefer it to be consolidated down to one kth component?

Participant: It’s hard to interpret this because nothing is highlighted,I don’t know what’s important here because everything is on there, so that’s in the text. I don’t know what I’m thinking is supposed to be important. The ellipsis suggests I don’t need to know how many there are there, but I don’t know. This diagram definitely needs text, or reading of the text, to know what the focus is.

Interviewer: Cool. Is there anything else you’d like to say about diagrams before we wrap up?

Participant: No. Diagrams are good!

# Participant 3 (P3)

Interviewer: How would you describe your experience level with NLP and CV?

Participant: Yes, Quite advanced I would say, with NLP. CV not that much, I just know the main concepts I’d say.

Interviewer: What’s your role at the moment?

Participant: I’m a Research Assistant at University of St Gallen in Switzerland, and a PhD student at University of Passau.

Interviewer: We’ll go through some specific questions, some examples, and then some card sorting. The first question is how do you use diagrams when communicating your research?

Participant: My work specifically doesn’t use a deep learning method, so I use diagrams [Pause]. There is no standard for diagrams in our area, so I try to make something that makes sense, or seems to make sense, and additionally put some explanation and hope the reader will understand.

Interviewer: Are you talking about conference proceedings?

Participant: Yes. If I want to make a paper and communicate my method using a diagram, I will try to make some squares with some colours and arrows, trying to make the most natural or intuitive diagram. I don’t know how to identify if it is intuitive or not, but I try to make some intuitive charts and explain it with additional text.

Interviewer: How do you decide what to put in the diagram?

Participant: [Pause] It depends on the complexity. Sometimes part of the method can be in pseudocode. To have an overview of the higher level of abstraction, I try to use a chart or diagram, and if there is something specific but it is still good to give the reader a clear idea of how it works I use pseudocode.

Interviewer: What tools do you use?

Participant: Mainly google draw, because it generates SVG, which looks much better in a pdf.

Interviewer: How do you decide what level of granularity to use in your diagram?

Participant: [Pause] Maybe the level of granularity is the same as the level of explanation of a concept, so to be more precise, when I am addressing my community I don’t need to define what e.g. semantic parsing means. If I’m talking about concepts which my community is already aware, that level of granularity is already enough. Everything specific to my work and part of my contribution I try to make that clear.

Interviewer: What kind of problems do you encounter when making diagrams?

Participant: As far as I know there is no standard, so it might be hard to understand, and maybe the most important point apart from your contribution, is to make your paper as clear as possible. I’d say maybe trying to make it clear, but sometimes I am afraid it is too basic, and that I’m making a silly diagram, that what I’m expressing is too basic. But on the other hand, it might not be clear for the reader. So this is the balance. If we had a standard that would be good, as that would say “at this level it is not necessary to communicate". Another point is about notation. Sometimes I try to use an object-oriented notation, when I want to describe an object and take a function. I have to assume people will look at it with this kind of mindset. We have two types of notation, object-oriented or maths notation, and in this case you usually work with sets and functions and defining tuples and so on, so sometimes it is difficult to define which one is better. I’d say using math notation sounds more formal, but on the other hand it is longer, and sometimes object oriented would be clearer to understand. But I don’t know if the reader would see it as sometimes bad if I use computer notation rather than math notation.

Interviewer: Could you clarify what you mean by objects?

Participant: Yeah, to give an example, let’s say we have a semantic parsing problem, and from one sentence I need to generate a function. My problem is I want to map natural language comments to functions, and during this process I need to identify from this sentence what is the main function description, and also what are the main object types, names and roles, and the parameter of the target function. In this case, it is natural for me to think of the utterance as an object, the text comment as an object, and after processing it would have some attributes such as function, potential parameter description and values, and so on. In this case I need to decide if I use an object-oriented view like C.verb, or set of objects with an index, like math notation. That would be the main point.

Interviewer: Are you aware of UML? What you described could be done using UML Class diagram?

Participant: Definitely yes, but the diagram we want to communicate is static. Usually the diagram wants to communicate more. The sequence of operations, the output, the execution and so on, so it is not a perfect match. Even if it was suitable, we have the same question of whether it will be seen with good eyes or it would be better to formalise using a mathematical notation.

Interviewer: How do you use diagrams when you are reading a paper?

Participant: I think it is the most important part of a paper, especially if it has a strong math behind it. I personally start with the diagram to get a general view and after that I look at the detail, and sometimes I can even skip the details if it isn’t very important for me. Sometimes if you have a sentence to describe the mathematics behind it, you don’t need to understand the details of the expression. It is just a formalisation and you already know what it means based on the type of algebra method used. For example, if someone says a process is made by backpropegation, you don’t need the formula, you already know what it means. Usually you can have this clear view by a diagram.

Interviewer: Let’s look at some examples now. If you could describe the diagram and what you like and don’t like about it.

Participant: [A Pause] I like the first one. [Pause]. I’ll try to be very critical, I probably wouldn’t care. In this after x1 and dense we have four circles, and it’s the same four circles for xk and xn, and all of those have different colours. Maybe it is not necessary to have different colours if it does not represent anything. Embed is yellow, which makes sense because you can see it is concatenated to the output of the other part of the network. The other colours don’t make sense.

Score is not clear, it would be good to have an explanation.Overall it looks quite nice, together with a few sentences beautifully clear.

Participant: The second one. [B Pause] I prefer the first format to the second. If I’ve understood correctly, we have some between the layers a change in the shape and something in the middle to share information, and a concat process at the end. It’s ok, it is possible to understand, but has less meaning. A high overhead to understand, and less meaning, compared to the first. If I’m not wrong, this is related to image processing, and as I’m not very familiar with it, particularly as this is the first impression.

Interviewer: Have you seen diagrams like this one before?

Participant: Probably not. Probably yes but I don’t remember. Usually I see things more related to the first diagram.

Participant: For C, [C Pause] maybe here those colourful images would be nice for someone who is familiar. In my case, I’m not really understanding them. [Reading labels]. It might be good for someone who can understand, because it seems to be a real output from the first image, so maybe it is better than the previous. At least it is more meaningful, the other is just many boxes and numbers to describe the shape. This one, if I’m correct that it means the real output, it is good, it is nice.

Participant: For D, [D Pause] I don’t like this one. It is not really understandable. Ok, some types come from the column… [Pause] Ah, there are the three colours in the legend, it is a bit more clear now. [Pause]. I can’t really understand what is going on. Maybe the words or chunks are associated with the encoder unit in the form of embedding. Hx might be a matrix, but then there’s the order… No, I don’t understand anything here to be honest.

Participant: For E, [E Pause] [Reading labels] This extraction model is quite complicated. It seems to have many layers. [Reading labels]. Having some scores calculated from the feature matrix which are concatenated to average pooling, and after that LSTM, and an encoder-decoder. I wouldn’t say this is bad, because probably it has a very complicated model. At least it tries to use something that the community are used to. The first part, input description and visuals are good, and described the input in a human-readable way. It’s using BERT here, and an attention mechanism, then it starts getting complicated, and the output is generally understandable. It is not clear by itself, the description being “the outline of our model”, but probably the text in the paper might describe in detail. It’s complicated, and not easy to understand. Even if it has a difficult architecture, the way the author has decided to describe the procedure extraction module in the yellow box is different from what the community is doing. It might be better to do it like they did for the green part. I don’t know.

Participant: For F, [F Pause] [Reading labels] That’s excellent. The model is easier, simpler. This is what I was talking about, concepts the community is already used to know. LSTM, you already know what it is, so a box is enough. Then input, projection layer, you already know what it is. Merge could be better explained. Fully connected layer is fine, but we don’t know how the input layer is because we don’t know how the merge was performed. Softmax and score also. Apart from merge it is very straightforward. That’s good.

Interviewer: Which one was your favourite?

Participant: [Pause] I think the first one.

Interviewer: Why is that?

Participant: [Pause]. I think that at the same time being high level it puts important detail. Maybe I like it because it sounds more formal, with variable names and math notation. Though it is not 100% clear. B is definitely the worst. [Laugh]. D is also bad. The others are ok.

Interviewer: For B you spoke about a lack of meaning. What are you trying to describe with that?

Participant: Lack of meaning is too strong, but sometimes we draw a diagram that is so easy to understand that we might not even need to have a diagram. Maybe B is this case. We have a diagram just to say the convolutional layers are changing the shape. On the other hand, as the chain is quite long it is good to have it visually. Text might be harder to follow. This diagram might be borderline on whether it is necessary or not. It isn’t lack of meaning, it is that the meaning could be replaced by just a few words, maybe.

Interviewer: The diagram seems to prioritise dimensional change with rectangles and numbers, in terms of its semantic content. Is it that you don’t value that semantic content?

Participant: Not exactly. Maybe this is the type of information that reading from a diagram is more effort than reading from a text. The main idea of creating a diagram is to simplify the understanding. If the concept is easy to describe in a few words, it is better to read a paragraph than to look at a picture and try to decode it.

Interviewer: How important are aesthetics of diagrams to you?

Participant: I think it is very important, especially as I don’t like overly colourful things. In the case of C I can understand it is useful to have a variety of colours, but if you can reduce the number of colours it is better. Even D uses just five colours, it’s ok, but it is good when the colour has a clear meaning. In E, there are too many colours, or at least shades of colour. It is unnecessary. The feature matrix has separate parts going to different parts, which looks good, maybe I am being too critical. F is nicer, just black.

Interviewer: Overall, how do you feel about the variety of diagrams used?

Participant: [Pause] It isn’t something I was paying attention to, but you are right [sic] there is a large variety. [Pause] Even when the diagram is bad, like D, which was difficult to understand, at least for me, and you read a couple of paragraphs, at the end of the day you can understand so you don’t tend to review in the way I am doing now. Naturally it is better if you can understand without reading part of the paper.

Interviewer: In terms of using examples in the diagrams, what do you think about that?

Participant: I think it is good, especially if it is complicated, because you can materialise. I guess this was the strategy of D, without much success, for C it might be good.

Interviewer: With D, as you say this centres around an example. What do you think is making this not work for you?

Participant: [Pause] Too many different notations. Compile rule and then Zr. Now I can understand, it might be a kind of grammar. [Pause]. Maybe I understand a bit better now, it is a set of rules from a grammar to parse the question to something. Because I saw a decoder I was expecting to have something related to neural networks, but no it seems to be a grammar or rule-based system, but it has memory and can select. [Pause]. Maybe this approach is too specific and not really common, so the author needed to create something that fitted and as we are not familiar with it we cannot understand easily from this information.

Interviewer: And the caption, did you see it?

Participant: Yes at first I think so, but let’s read again. [Pause]. It doesn’t help that much, but it complicated my life more because at first I thought it was neural, then not, and now it says it is. And the author’s caption describes in an easier to understand way [Reads out]. The complicated part is not described and the easier part is described.

Interviewer: Anything else about those examples?

Participant: No. Maybe the caption. I personally like to write large captions, and sometimes the reviewer asks me to remove the caption and put in text. But you know, I like to put large captions, because imagine someone who wants to understand the problem or the approach just from the diagram. If there is something complicated or confusing you can describe there, and the caption is a natural place to search for explanation, and if you have to rely on the text you have to search for a mention, and if it is printed out I can’t do ctrl-F command. I like to have large captions, and most of them here are small, half of them are just useless. It’s just something to say the minimum.

Interviewer: Moving on to the interview task cards. Pick out the top 2 or 3, and any other things, including those you definitely don’t do.

Participant: [Reading]. What do you mean by opportunities to alter the architecture?

Interviewer: Looking at it and saying “oh maybe if I put attention in here I could get a new paper out of this”

Participant: Ok, I understand. [Reading aloud] [Pause] I think all of them are quite important. Less important is “Identifying what the author thinks is important to communicate”. Maybe it should be the same as system novelty. I know sometimes we are not good at communication, but I’m not interested in the author. “Index to navigate” is also secondary, and maybe possible to get from the other parts. Top three: “Understanding how the system works”, “writing summary” definitely especially now I’m doing related work of my thesis, “identify novelty or contribution”. I think these three, but the others are also important. It is good to see external dependencies, corpora and data types, layers and relationships, everything is interesting. Maybe “altering the architecture” not now, I’m not reading with this mindset.

Interviewer: Is there anything missed off this list that you think is important?

Participant: No I don’t think so, because “how the system works” is almost everything!

Interviewer: The last part is a “have we missed anything”. Do you use diagrams to reason as part of your creative process?

Participant: [Pause] Good question. When I’m alone, I usually. When interacting or doing something collaborative, it is better to communicate and understand the ideas [through diagrams]. I drew some diagrams when sketching my thesis, just simple ones. It is important to use diagrams to help with the thinking process, I think so. I have some boxes, arrows and titles, to try to make sense of them. I wrote diagrams to explain my method in papers, and now I use it when writing my thesis. I have a section or subsection that might match different topics, and try to name those blocks of content to see how they best fit. Sometimes we can pivot the content, grouping related work by method or by task. Should sections represent the tasks or the methods? And then I need to regroup, so sometimes it is good to have diagrams to decide about whether it would be better to organise in such a way that sections have more or less the same length. It helps with these things.

Interviewer: You were talking a lot with your arms then, almost doing a pivot with your arms. Do you identify as a visual thinker?

Participant: [Pause] I think so. I’d never thought about this question, but I think so.

Interviewer: Alternatives are audio, or kinesthetic so like doing…

Participant: Audio definitely no, I prefer to see things.[Aside topic about a related event in the gym].

Interviewer: Can you think of any tools that would help you make or consume diagrams?

Participant: [Pause] Yeah, as I told you, I use google drawing, but it isn’t really made for diagrams so sometimes you have to adapt stuff. To replicate part of diagram you have to take care to group the objects that you copy. It would be nice to have something designed for diagrams. Maybe Microsoft have but I don’t know, I’m a linux user. I also tried to use inkscape, because the main idea is to generate an SVG, but in this case inkscape is even lower level, so it allows us to do more but for more effort. More control, but you need to do everything by yourself. Generally speaking, I think google drawing is quite good. In the case that someone, maybe you, come up with a standard, it would be good to have a tool. For example how LSTM should look like, and the type of merge let’s say concatenation, and you just link it automatically. That would definitely be very nice.

Interviewer: Anything else?

Participant: No, everything is there.

# Participant 4 (P4)

Interviewer: What is your experience level with CV and NLP?

Participant: In the past I worked with NLP but now for my PhD I work with image processing, more focused on machine learning and convolutional neural networks. Basically my PhD has the proposal to take an image and describe it in a semantic graph of interlinked concepts. [Sentence redacted for anonymity].

Interviewer: So you have an NLP background and now are doing CV. Which one is stronger for you?

Participant: Maybe both because I use both, but maybe image processing because I’m applying that now.

Interviewer: Could you describe how you use diagrams when communicating your research?

Participant: I don’t have any specific rule because I’m trying to express information to the reader of the paper, so I am agnostic about diagrams. I choose the best way to transcribe my ideas to the paper. There is no specific rule.

Interviewer: With what frequency do you use diagrams? If you wrote 10 papers, how many would you guess have diagrams in them?

Participant: Probably 10. It is not easy to explain a complex network without a diagram.

Interviewer: So you use the diagram to describe the system architecture?

Participant: Yeah, in many levels. Sometimes I want to describe a specific part of the network. Sometimes I want to group part of the network. For example, I don’t want to describe LSTM because it is common sense for the reader. Sometimes I want to specific characteristics of the network, such as parameters, optimisations, or specific things which will impact the result. I think there are multiple levels which you can express yourself, and you include what kind of information you want to transcribe to the reader.

Interviewer: How do you decide what kind of information to include?

Participant: [Laugh] I think I try one thing, my supervisor reads [Laugh] and they don’t understand what I’m trying to say, so I try to change some things. I try to test with my supervisor or some friends before I try to publish the paper. To me there is no specific rule to decide what information to include.

Interviewer: Are you doing that in one diagram, or do you draw several?

Participant: Usually several

Interviewer: And you iterate through?

Participant: Yeah

Interviewer: How do you create these diagrams?

Participant: I try not to start from scratch. I get some papers or tutorials, and try to see what kind of diagrams people usually use, and sometimes I try to combine them for my needs.

Interviewer: Is that happening in software or on paper?

Participant: Usually I use draw.io, if that is enough. If I need a proper diagram, I use vector tool on linux, called inkscape.

Interviewer: Do you find you are using some elements repeatedly?

Participant: I try to describe the unit that will repeat. For example LSTM, suppose I was publishing something about it, in the first moment I would describe “ok, this block is the LSTM” and we have these nodes and they are connected. In a second moment, using a combined LSTM for some specific task, I just use a block with one name. The same for CNNs and ResNets, I try to explain the unit of residual that I am using, and after just use the blocks concatenated to show how my network is.

Interviewer: How do you decide whether to use a diagram or text, or equation, or pseudocode, or something?

Participant: Usually I prefer diagrams. It is a more philosophical question, but I don’t like the way papers are created, because there is too much text and it takes a long time to read. Usually I prefer to use a graphical representation because it reduces the time to understand the idea that is being expressed. Text for me is only to explain something that is not easy to do with a diagram by itself. If possible, we would change the way completely that we write papers, it is too old for the 21st century. We could use more things, more graphical, hypertext, or something like that. My first option is always diagram.

Interviewer: How do you use diagrams when reading a paper?

Participant: When a paper has no diagrams, I try to draw it myself, because to me it is easier to understand visually than trying to interpret a block of text. When the paper has diagrams which are confused, which isn’t unusual, I try to transcribe the diagram to express what at that moment I am trying to understand. For example, now I am working with attention models, I have a set of things which are important to me. I read tonnes of papers, and I try to summarise each paper in a diagram, the same kind of diagrams, to understand well the intention of each author.

Interviewer: Do you see any problems with diagrams that you see?

Participant: Yes. We do not have a formal language to represent things. It takes a long time to understand what each one wants to represent, mainly because many times (it is not unusual), people put symbols like triangles, squares and don’t explain the meaning of each one. If you had a more formal and easy language, not UML because this is too complicated for me, but something like that with a specific semantic, maybe it would be easier to communicate.

Interviewer: Shall we move on to the diagram examples?

Participant: [Misunderstood question] There is a guy in udemy, lazyprogrammer, who has a lot of interesting classes about ML, and he uses tonnes of diagrams there, and has a lot of good representations of neural networks. At least it helped me when I was trying to understand CNNs, mainly because the representational change is always represented in 2 dimensional operations, but in truth you have a cube when you are making a convolution, and it took me a long time to understand it. To translate a convolutional such as a network sharing weights between neurons, this guy has a good representation. I don’t have a general good diagram, sometimes I read things on Medium or in papers and I like it, I get it, it is easy for me to use. I don’t have a specific methodology to do this.

Interviewer: Shall we move on to the diagram examples?

Participant: [A Pause] A shows largely layers represented by Dense. I don’t know what dense means here, one neuron or a network I don’t know, because you have x1, xk, xn, this part is confusing. It is not explicit what is core in the diagram, I prefer if you define what type of function or parameter you are using. I like that there is softmax here, I think this is a classification problem because usually softmax is widely used. The Genre-aware attention layer doesn’t mean anything to me. And the symbol in the right bottom for embedding layer, I think this is a matrix of data, but it is not explicit and doesn’t tell me much. That’s it.

Interviewer: What about “score”?

Participant: I don’t know what it means here. It is a score function to measure the accuracy or a function, but I prefer to state explicitly what type of score is being used, for me it is not clear.

Interviewer: Overall what is your feeling about “A”?

Participant: It’s acceptable, I see a lot of diagrams like that. It isn’t the worst, nor the best. Maybe I could work with this diagram for a paper, but with more information and with a lot of text to support the description of the architecture of this network.

Interviewer: B?

Participant: [B Pause] This is the most classical diagram! It’s a CNN, I can see because for me this operation is common. Convolutional with cubes. There is not much here, for example I think there is a [maxpool] or some kind of operation because they are reducing the dimensionality, but I’m not sure. No, there are just convolutional operations, I had to increase the size of the diagram. Ok now I can see. Third person… The inputs are not clear, they look like an image and they put the bottom side I don’t know what that means. There is a sequence of convolutional operations, probably a segmentation. I don’t know, the task is not clear. Now I read the Siamese network and now I know the purpose of this diagram [Laugh]. They are sharing, I don’t know what they’re sharing, the weights or anyway. I know because I work with siamese networks, but in the first moment it is confusing to understand what is being shared. It is a more classical sort, and easier for me to read this kind of diagram. In the end we have the embedding space, so I know the purpose of this, and we have a contrastive loss. It’s ok for me, this kind of diagram is common, I can read it well, I can understand what is being represented. There is not much information about parameters, convolution functions, if there is something like back normalisation applied, but it’s ok, I think that is a good diagram.

Interviewer: You’ve seen ones like this before. Do you think it is because you have seen it before that it is easy to understand, or is it intrinsically easy to understand?

Participant: Both. Because this kind of representation has cubes in sequence, which is really common to work with convolutional neural networks, but to understand the whole of the diagram I only understand because I worked previously with a siamese network.

Interviewer: Do you think they need to be 3D cubes? I’m just curious.

Participant: I think it is, but now in the moment 3D doesn’t give me much information. For me it is important, I have [3D] matrix input image, a channel with convolutional operations, so I will have one layer with 64 channels and I apply the second convolutional so 3D or 2D to me here presenting the operation with the number of channels would be enough. 3D doesn’t give me much information.

Interviewer: The numbers, are they important to you as part of this diagram?

Participant: Totally, totally. Very important. For example, in the first operation I know it will have 64 convolutional operations, each one creating one of the channels, so I can understand with these numbers.

Interviewer: Is there anything you don’t like about B?

Participant: Probably there is a lot of information which is not represented here, such as specific operations like back normalisation or max pooling, the parameters of the network are not represented here, so I don’t like the loss of information with this diagram. I don’t know, I like it.

Interviewer: C?

Participant: [C Pause] Oh C is really noisy, I don’t like it. I don’t like this kind of diagram. [Pause]. Confusing. I don’t understand what is happening here. No I can’t understand this, probably I would take a long time to understand what is happening here with this architecture. I know that I have an image, I will apply a pretrained guided net that will generate the output Dc. Guided disparity net. And after, I don’t know what these fused clouds will be generated and apply the projection in a way combined with the guided disparity map. I don’t know, I cannot understand at all well this diagram. [Pause]

Interviewer: Can you see it has an example going through it? It is the same example image which it is centred around.

Participant: [Pause] I don’t know if I’m understanding the question. In this case, there is the image of the car as an input. I understand it is possible to [use an example to] make some kind of interpretation of the image through the steps of the network, but usually it doesn’t bring much to me. For example, you have the image which is a car on the street, and the second image, I know this came from the previous image but what kind of information is represented in here. I know this is a “guided disparity map” but I don’t know what that is, so this kind of information doesn’t mean much to me.

Interviewer: Is there anything you like about this diagram?

Participant: I think there is nothing! Sorry but I can’t say anything positive here. [Laugh]

Interviewer: [Laugh] There are a variety of different diagrams here. D.

Participant: [D Pause] We have an encoder, I understand what an encoder is. Memory. Title. [Reading labels]. I think this diagram tries to transmit the idea of a sequence of operations, because we have the arrows, we have boxes, but again in the first moment I cannot understand well the information that is trying to be transmitted here. I like the idea of a sequence, I like maybe because I have a background with NLP, but I don’t know what kind of operations are made in each box. My first interpretation would be this. I don’t know what means memory and schema in this. I have a background in NLP, I understand these concepts, but I don’t know how they are used here. Here we have the chart with two bars and the second one looks like it is bigger, perhaps transmitting the idea that year is more important for the sentence than book title, I don’t know. I know that’s memory, but I’m trying to interpret it like something I’ve never seen before. I think that’s it.

Interviewer: What do you like about this?

Participant: I like this diagram, but again I need more information related to the sequence of steps and how they are being applied to the text. [Pause]

Interviewer: And would you expect to find that information in the text?

Participant: Maybe instead of one diagram, there could be a sequence of diagrams showing how the state of the network is. Suppose we process the word “show” first, ok, now my initial state is this, I will apply the word “show”, and “show” will change the shape of my architecture in this way. Now in a second diagram I will apply the input “the” and the state will change in the way. To show a sequence in a stacked image is not an easy task. I prefer to use a sequence of images to show step by step what is happening to process the text.

Interviewer: So you’re saying it might be more effective to pull it apart so not all the example phrase is being shown simultaneously?

Participant: Exactly. Exactly. To me, CNN is about structure, so you can show the whole. But to NLP where you talk about Recurrent neural networks it is about the sequence of things. It is not easy to show the information of steps in a sequence. I prefer in this case a sequence of multiple images describing what is happening in each step in the life of the NN to process the data.

Interviewer: What do you think of the use of colour in this diagram?

Participant: Encoder, decoder. Ah, now I see the legend, the embedding encoder and decoder units. I would have to read the meaning of each colour.

Interviewer: Do you think it needs colour?

Participant: Yeah, to show what is encoder and decoder, and the same for embedding. How this guy could use the same visual representation without colour, I don’t know. [unintelligible]

Interviewer: Anything else?

Participant: I like the schema encoder present in the diagram. It is not too expressive. I like but at the same time I don’t like. The second one, I don’t have anything I really like here.

Interviewer: E?

Participant: [E Pause] Ok. I read this paper, I’m sorry! I’m compromised. I know I have input video, it’s good. Showing the idea of input sequence. We have resnet-34 here, ok. To transform this input video into something. [note, the participant went left to right rather than following arrows, completing the visually unobvious transition of the video to fusion module]. I’ll try to interpret this diagram without using my previous knowledge about it.

Interviewer: No that’s ok, you do whatever is natural.

Participant: Ok. The self-attention layer in the transcription embedding is clear, so I like. It is a complex diagram because they are using image and text together. The information is combined with the information transcripted from the image, in a network. The result of the context aware fusion module is used for the procedure extraction module. This box has a feature matrix, I know what a feature matrix is, so this is easy to understand. The second box is not easy to understand, the right side of the feature matrix, but I know what it is, so ok. We apply LSTM for procedure prediction, with a start and end state of course. Information will be sent to a procedure capture module. Then after will apply encoder decoder, to get the mixture in a bowl. This diagram is too complex, use image and text, using resnet. Maybe this is easy to understand because I know what resnet is. Uses LSTM, attention modules. This information is combined and in the end it generates “spread the mixture and potatoes in a bowl”. I like this diagram, it has too much information but at the same time it becomes clear. I know what is happening here.

Interviewer: Is there anything you don’t like about it?

Participant: The procedure extraction is confusing. The step from feature matrix to the second box. I know the output of resent-34, but of course resnet is a network by itself, with a complexity that is not present here. At the same time, it is too[sic] representative, and a lot of information is expressed, but I don’t know a better way to represent this complexity here. So, maybe more information but I don’t know how. I would need to think more to improve this diagram, it is really good.

Interviewer: I’d like to explore some of this with you because you have a depth of knowledge. In the CAFM there is an iconic representation of 10 rows with circles and rectangles. Is 10 a special number do you think?

Participant: I don’t know the meaning here. Maybe 10 frames from the image, I don’t know, I don’t remember.

Interviewer: In each of the FM, these are arbitrarily sized. Is that a problem or is that fine? Should they be smaller?

Participant: I don’t know. If 10 is important, it is ok. For the feature matrix, 6 columns and 13 rows, I don’t know. If this information is not useful, of course it could be smaller, or a smaller number of points in CAFM, would transmit the same information to me. If 10 is important to the architecture it’s good, but if not then the size of the matrix doesn’t matter. Except if you have one row and one column, it becomes a point. And the same if 1000’s of rows and columns.

Interviewer: On this “arbitrary” topic, in the transcript embedding, self-attention layers are missing. Is that effective?

Participant: I think it is. Of course in these diagrams you expect a lot of information, so it’s ok, it isn’t a problem. You take the output of BERT, and apply the layer n times.

Interviewer: What I’m trying to explore is the arbitrary dimension, they’ve omitted and left out the transcript embedding, but in CAFM they include 10, for example.

Participant: Ah yeah, that is not confusing to me, but you’re right. It’s not a problem for me. If you keep the number well defined, for me it is better, but it’s not a problem.

Interviewer: Anything else about this diagram?

Participant: No.

Interviewer: Ok, last one. F.

Participant: [F Pause] Automatic learner summary assessment for reading comprehension. This network is a summary, so I imagine this network is applied to text. Apply a complete text, projection layer, apply bidirectional LSTM, mean over time, mean over time, merge operation will be applied, but I don’t know how, so the first question is here. This merge will generate an output which will apply to a FC layer, and of course we apply a softmax score, and probably classification or that kind of operation because we’re seeing softmax. Ok. This kind of diagram is too simple. It transmits information but a lot of information is suppressed. Now I have more questions about what is happening here than information about the architecture.

Interviewer: Is it specifically about the merge?

Participant: Not only, because what kind of projection layer, what projection type is applied here. How is this projection used in LSTM, I don’t know what the projection is. After, mean over time. I imagine time here means the word after. But maybe something about a slider, if there is a window of information that is being processed such as BERT uses. And after, for the merge I don’t know what kind of merge operation, is that concatenating vectors, are there vectors, I don’t know. And I don’t know the output. The output is a score but what kind of score, is it a probability or I don’t know. I don’t know what kind of output is here.

Interviewer: What do you like about this diagram?

Participant: I like the simplicity, but at the same time that is the problem. At first glance I can understand the parts of the architecture which are here. Maybe I would use in a first moment, [unintelligible 42:15] but after I would try to explain without the box, step by step. Maybe showing overview of my architecture is this but now I will show you the projection layer and how it works, and after the mean over time operation and the merge. I like this as a summary of the architecture, but I don’t like if it is the only one diagram present. Then it should explain the whole of the network.

Interviewer: Anything else?

Participant: No.

Interviewer: Perhaps we can move on to the task list. This is a list of things you might be doing with diagrams when you are reading them, and it would be good if you can pick out the top 2 or 3 things that are most important to you, and any things which are not important at all.

Participant: The first one, the novelty of the paper, it is important, but not the most important thing to me. Because probably the presentation cannot show what the author has presented as the novelty. Identifying dependencies between layers, is really really important. External dependencies is not so important. Corpora and data types is important. Representational choice is important. Purpose of system, no. I hope before I see the diagram I have previous knowledge about what kind of tasks are being computed.Specific features, sometimes important sometimes not, so I’d say it is important. Opportunities to alter. No, that is not important to me. Author thinks is important, this is important. Comparing to other systems, we don’t have a unique language so it is not important to compare in that way. So, to compare to background work. Initial check, no. Index to navigate, no. Memory aid, no. Aid for summary, maybe yeah it’s important. How the system works, yeah it is important.

Interviewer: Which are your most important three?

Participant: Most important is identifying Layers and internal dependencies. The last one, how the system works, is important. And those two are the most important here.

Interviewer: You’ve already mentioned some you don’t think are important, so that’s great. Is there anything missing from this list that you think you use diagrams for?

Participant: Maybe because now it is a pain for me, but identifying the parameters of the architecture, because when you try to reproduce the paper, I usually take a long time to discover each parameter and will not talk about how to optimise them, but what kind of parameters are present in the architecture. I think that is necessary to represent.

Interviewer: And you feel a diagram is the right place for that, rather than in a table or text?

Participant: Maybe a complementary explanation in a table, but maybe some information such as the number of convolutionals, there’s that. You asked me about the numbers before, and I use these parameters, they’re important. Maybe not all, but some of the parameters should be present in the diagram.

Interviewer: The last part is just some more general questions about diagrams. Do you use diagrams to reason as part of your creative process when doing research?

Participant: Yes. Yes. Yes. As I said, I think in more visual ways, so in my creative process, I need to draw it, I need to write something in the paper to not only imagine but to think well about the paper, about sequence, about its choice, so it is totally important to me.

Interviewer: So this is to help you understand what is happening?

Participant: Yeah. And two things. OK, if I give an example, now I am working on a problem with a small number of data and I know I cannot use a really deep network, except with transfer learning. To understand the problem, I don’t have much data, I need to create something that is deep but not so deep. I know that if the data has lots of noise what kind of operation I will have to apply such as back normalisation or something like that. So this helps me.

Interviewer: Do you draw diagrams then before you are building the system?

Participant: Before, during and after [Laugh]. The whole process.

Interviewer: Do you do those online or on paper?

Participant: In the first moment on paper, but how usually I work with my friends is I output on a board and we discuss the options. So, in the first moment on paper and after on a board.

Interviewer: You’ve sort of answered this already but I’ll ask in case there is anything else. Do you use diagrams in any other way during your research more broadly?

Participant: yeah yeah, naturally to show the ML architecture, but when working with systems that use machine learning as a part of them, I need to show the whole system and of course subsystems and specific points where machine learning applies. I use diagrams to describe the whole system, to describe the steps of the system, to describe the sequence, to describe the components and how they are interlinked, to describe ML architectures, and so on.

Interviewer: Wow, that’s like everything!

Participant: Yes, I’m addicted to diagrams.

Interviewer: Is that because you are a very visual person?

Participant: Yeah yeah, I have a friend that is a math guy, and he puts everything as formulas, and he works well, only looking at formulas. But for me, it isn’t easy to understand him, so we communicate with each other trying to translate what we draw into formulas and what he writes in formulas to diagrams.

Interviewer: Amazing!

Participant: [Laugh] I don’t think it is amazing, it is necessary!

Interviewer: [Laugh] Can you think of any tools that would help you in making or reading diagrams?

Participant: I don’t know why there is not yet something that lets me not only draw my architecture but also produce the code of this diagram. I know there are some of them, but for example, if you get keras from using tensorflow and pytorch, you have a well defined components there. I don’t know why you can’t use something like a drag-and-drop and take these diagrams to code. So I don’t know if it is the context of this work that you are working on, but I am missing something that helps me to test me ideas. I spend a lot of time translating from English to code.

Interviewer: Have you seen tensorboard in tensorflow, it is a graphical interface?

Participant: Yeah, tensorboard is too confusing. I just use it to see the loss functions and how my networks are behaving. But the graph is completely crazy, I don’t understand well what is happening.

Interviewer: Is it too granular?

Participant: It is too granular. I came from a business intelligence background, and they have the concepts of drill-down and drill-through, sometimes you can see things more simplified, and sometimes you want to focus on specific parts, and this avoids showing me specific things and presenting too much information to be able to see in a first moment.

Interviewer: Fair enough. Is there anything else you’d like to say about diagrams before we finish?

Participant: We have to unify the language of diagrams because I take a long time understanding the visual language of each author. But I don’t have much more to add to this interview.

# Participant 5 (P5)

Interviewer: How would you describe your experience level with CV and NLP?

Participant: I’m a practitioner, I would say. Mostly NLP, less so CV, although I probably would be able to understand what they are talking about in a paper, to a certain degree.

Interviewer: And you’re in your final year of your PhD?

Participant: Yes

Interviewer: Cool. Can you describe how you use diagrams when communicating your research?

Participant: Well, honestly I don’t, because my research is less about [Pause]. I mostly use tables and figures. The last paper I wrote didn’t have any diagrams, because the data was mostly textual and since I’m working in linguistic features it is very hard to express them in diagrams. The things I used were colour highlighting and text, but no diagram. Although I did have a schema, drawn up as a tree, you could call that a diagram I guess.

Interviewer: You make systems as part of your research?

Participant: I did some. And yes for that we did use diagrams to explain it.

Interviewer: In the case where you were making system diagrams, how did you decide what to put in the diagram?

Participant: [Pause] For the one where we used something similar to [redacted], we went with what people in the literature were using. The people we cited; their diagrams. For the other one I didn’t have any considerations, just something that described the system best from my point of view. It is mostly data flow guided. Like a sequence diagram, where one dimension is the time, I try to do that, where time is not a dimension but an arrow through the diagram. Where you have an input datum or an input action, to what data processing or following actions happen, and there is one red line which is basically this intent or action or datum and you just follow it and see what happens to it.

Interviewer: So that wasn’t a NN diagram?

Participant: No

Interviewer: Was it fairly standard for the field?

Participant: I don’t think there is a standard for the field. One of them was HCI and one of them was system description, I don’t know. I could probably have used UML but UML is only for code while this was for the whole concept, like architecture. I don’t know. Any time I’ve drawn an architectural diagram I haven’t followed a standard because I am not aware of one standard way of drawing architecture diagrams.

Interviewer: When you are drawing an architecture, how do you decide what to include? At what level to abstract, how to group, and so on.

Participant: Basically in a way that supports the claims in the paper. I don’t think anyone reading the paper is interested in which variable name you use for this class or whatever, but more when you look over the picture for 30 seconds or 1 minute, that it supports the case that you are describing in text. Usually I try to describe a usage scenario, like in HCI, give some NL input and a bunch of things happen to it. The bunch of things is from the user’s perspective, but if you want to describe how it works that is from the system’s perspective, such as there is POS tagging, and some DB lookup, and the user gets presented with five choices or whatever.

Interviewer: So it is about constructing something that supports the claim?

Participant: Yes, supports and elaborates. Text is linear, while the figure isn’t. When you want to say “3 things happen at the same time” in text you have to enumerate them, while in a figure you can just draw 3 arrows and it is better visible I guess.

Interviewer: You’re leveraging the 2d nature of the diagram to communicate more visually?

Participant: Yeah that and… text is linear so when we say “and then we return back to step 2” we need to backtrack. One way of expressing that would be an algorithm, but reading algorithms is probably not as easy as reading a diagram. Algorithms also require a different level of formalism.

Interviewer: What tools did you use when making your diagrams?

Participant: yEd, Tix. That’s it. Mostly yEd because I know it. I am aware that noone else uses it, and it makes funny looking diagrams.

Interviewer: What sort of problems do you encounter when making diagrams?

Participant: [Pause] There is no standard. If you are formalising an algorithm, you can write pseudocode. That doesn’t specify granularity, but at least there are formal concepts like for, while, and so on. There is a finite set of things you can use to describe an arbitrary formalism. Well not arbitrary, but falling in the hierarchy of computability and everything. With diagrams you can basically do “whatever”, and while you are doing “whatever” it can either be good or it can be crappy. You have more choices to choose from, and more opportunities to end up with something that doesn’t make sense. You look at a diagram you’ve just made and are like “this doesn’t make sense” and start from scratch, or start moving things. Every time the reviews come back, “oh this diagram is crap, can you change it”, and then the next review “oh this diagram is actually this”... [Pause]

Interviewer: Do you find reviewers are commenting on the diagrams?

Participant: Sometimes they are, and sometimes inconsistently so. I got a couple of times “this diagram is not clear, this diagram is overloaded, there’s no spatial consistency in it, or blah blah blah”, especially if you submit to HCI venues I guess, there are people paying more attention to it. From other venues I got more like “I don’t understand this” but not why or what’s wrong.

Interviewer: Moving on to how you use diagrams when consuming research

Participant: Right, so I tend to inductively understand something. That is, from an example, generalise to how it works. Diagrams usually come with an example, like for a neural network or architecture or model diagram , if they are trying to describe what happens with the vectors they will use some words or some pictures, some vector magic happens. That is easier from a diagram than from the corresponding text or mathematical notation, where they say we project this into that space. When you read the text and it is accompanied by a diagram, you see the whole process at once. I guess it is like an anchor to where you are in the next right now, so you can correlate between what they are doing and part of the diagram, and I guess it is easier to remember or to “save in your brain” the diagram than the textual description.

Interviewer: Are you using it almost as a summary of the paper?

Participant: No, not of the paper. The things that matter are hardly ever put in diagrams, even if people writing the papers are trying to do differently. It isn’t the architectural novelty that is interesting, it is that using this architectural novelty you can obtain these results which are summarised in tables, not in diagrams, unless you draw some charts. Charts or tables are effectively the same. About these architectural diagrams, it is not a summary of the paper, it is a summary of the methodology or the approach, like a blueprint or car schematic.

Interviewer: So if you had a paper in front of you right now, how would you interact with that, do you go straight to the diagram?

Participant: No, I read the abstract, then I decide if it is interesting for me or not. If I decide it is interesting, I decide which part is interesting. Let’s say a typical paper has 4 parts, Introduction, Literature Overview, Methodology, and Evaluation. Not all papers have this, but let’s say it is one where they are using a neural network. Depending on what I want to know, if I wanted to know the data they were using or the evaluation, I would read the evaluation. If I’m interested in the architectural novelty of that I would then go the methodology. If I wanted to read the paper properly I’d probably read the introduction first, as it gives a cognitive map as to how to read the paper. Then I’d try to understand from the paper. If that doesn’t work out, I’d usually go to the internet and try to find a blog post that explains that in more depth. Like the Transformer paper, it is super well explained in the paper but still noone understands it so there are a lot of blog posts describing that paper in greater detail.

Interviewer: So the diagram is used if you come across it at some point?

Participant: No, so the diagram which I would assume is architectural, when I am reading the methodology part I would use the diagram for a better understanding. So again, I try to instantiate an example, usually the example on the diagram, and try to generalise from that. They usually don’t put examples in text, the example is usually in the diagram, and from that example it is easier to understand what is going on.

Interviewer: So is it that you are trying to find the examples, or is it that the diagrammatic representation is particularly useful for doing the aggregation and making it so you can see everything in one view.

Participant: [Pause] I think it is both. A start to end example also helps to give [Pause] a summary view. But not in terms of how it performs all together but how it performs start to end, I guess. Thinking of the Transformer, you have one word vector Cat and it goes through a bunch of attention vectors and goes through some other vector in some latent vector space and is decoded. How it goes through the attention layers and where the keys are declared and multiplied, seeing that in a diagram gives a summary of how one vector, one word, runs through the whole system and gets spat out.

Interviewer: Cool. What problems are there in consuming research more broadly, in CS?

Participant: You get a paper which describes everything quite well, then you try to replicate it and you find it doesn’t describe it all that well, there is a lot of things missing. If you are lucky there is some code associated with it, and you run the code and see it is doing something completely different. It is doing the same, but the way you describe it in code is very different, especially for these neural network things. Like, you don’t necessarily see this is a projection to a lower dimensional vector space, what you see is a linear layer, another linear layer, a matrix multiplication, and to correlate that is not always necessarily easy. At least for me, I’m not a mathematician. I studied some linear algebra and analysis at undergrad but I am not an expert on it. I can’t look at a diagram and know exactly what mathematical operations are behind it. Or I read through text that says “we’re doing a projection here” and understand it means we are doing this, this and that, and that in Torch or Tensorflow it means we need to do this, this and that operation, and it goes without saying that when you use these parameters with this this and that distribution. These things are not necessarily reported, and it is not evident it should be like that. If you are lucky it is hidden somewhere buried in code somewhere, but often it is not. On a broader topic, there have been things discussed about reproducibility, about choice of hyperparameters, there is no principled methodology to do so. Even if you run the same code, or if you tweak it slightly, you might get vastly different results. There are people more acquainted with that, proposing methodologies to get a more principled approach there. I’ll be happy to see how that pans out, but in general these are the issues. Also, by just reading the paper you don’t get a feel on how much the evaluation tells you about how I can apply it in a slightly similar or a slightly different field.

Interviewer: On to some examples. There are 6 examples here, what would be great is if you can look at the diagram and tell me what you like and don’t like about it.

Participant: [A Pause] Ok so that is some kind of text classification task, that uses some sort of Genre information as additional input, and they compute attention score between inputs, I assume words, with the Genre which I assume is a label. Then they concatenate it, sum up all the inputs, and classify to whatever extent. Either sentiment or whatever. So yeah. I can only say this much without the corresponding paper, but I already mentioned that. I can see how… This one doesn’t have an example with it, so I can only guess what genre, x1 and xk is. It would be better to have instantiations. I think it would be better. [Pause] Yeah also this is conceptually the same thing [the vector circles] so why are they different colours. The other parts are the same thing in the same colour. [Pause] As far as I understand, they want to sell this genre-aware attention. I don’t know, I’m not too sold, but I’d probably need to read the paper. Some of these choices seem a bit odd, for example colours, and why the embedding layer and the dense vector. Dense, what is that. What kind of vector is that embedding. Why does this have embedding and why doesn’t the other part. Looking in the paper I would probably find answers, but just looking at the diagram itself. It also bugs me that the embedding is not symmetrical. It’s not aligned, and not symmetrical.

Interviewer: Is there anything you like about this diagram?

Participant: Oh yeah, I guess. I guess the guesses I made were correct. I can see what they are doing. I don’t necessarily see how or on what data. But I can definitely see what they are doing, they’re doing some text classification using additional genre information, and using some attention. Most of that is coming from the labels of the diagram [Laugh] I would not necessarily be able to tell that this is attention. But I think the labels are good. At least it helps here to understand this is an attention layer. But then again knowing how attention works, if they had just put attention here I would already know, it is implied that this is happening, because that is how attention works. If I didn’t already know, I probably wouldn’t be able to tell how it works just from this diagram.

Interviewer: Just a box that says attention would be sufficient to get the same information?

Participant: Yes. Well, [Pause]. Yes. From this, I can’t tell what is attending to what. Are you attending hidden vector as a whole to the whole embedding and you need to do a cross product to get the score? The score is between what, right. Is it each dimension of the vector here and here. Is it just a dot product between both vectors, and you multiply that vector with the whole vector, or do you multiply the whole of that with that vector. At that point, [of ambiguity] you might as well just say attention. And then explain it somewhere else. Yeah it is attention, but what attends to what I wouldn’t be able to tell. I can guess, probably they are doing a dot product between this and that, and then some learned score and multiply that with the hidden vector. But also it might be completely different.

Participant: [B Pause] Where did you find that? Right, so that’s computer vision. I am not sure if a 2 stream FCN is something that people in that research area are aware of, or if that is what they are proposing. Again maybe this is common vocabulary, like attention in NLP where everybody should know about it. [Pause] They’re embedding some pictures, and computing a loss. [Pause] To understand that I definitely need the paper. I understand the architecture but it might be doing anything. As far as I understand they want to identify common people across different videos. [Pause] or frames, like I don’t know what the input is, probably video frames. So given two video frames you want to say or segment out people that are common. I don’t know the usual method or the bread and butter for segmentation or bounding boxes, I know it exists but I don’t know how the internals work. Therefore this diagram doesn’t tell me too much, but it probably because I am not an expert in CV. [Pause]. The overarching idea with the loss is probably to have the segments of the frame closer together, and ones that aren’t further apart. [Pause]. Yeah. [Pause]. Are they multiplying this with something? I think it is multiplication, or does it produce this? I don’t know.

Interviewer: There are a couple things you said there, one was about outcome, what the purpose of the system is.

Participant: That I got from the caption. Without the caption this would say literally nothing to me. Not literally, but figuratively nothing.

Interviewer: So what are the numbers?

Participant: Right, I guess that is [Pause]. It is 4, the number of channels, and 3 also the number of channels. I guess colour image and greyscale image. Then since the convolutions are 3d I guess the width and height are given by the dataset, and the dimension is of the convolution layers or the linear layers in the fully connected one. I guess it is the hidden dimension of those layers.

Interviewer: The diagram is communicating to you it is using convolutional layers by these little arrows, is that right?

Participant: I would assume so.

Interviewer: If it didn’t have the arrows with labels on, would you know it was a CNN?

Participant: I would guess it was a CNN, as that is what tends to be used for image tasks.

Interviewer: What are your thoughts on this contrastive loss representation?

Participant: I don’t know what contrastive loss is, so I can’t tell. I think it is supposed to move points closer in the space that are conceptually similar, that is they have the same features, and move points further apart that don’t. That is what I would assume, but that is because it makes somehow sense to me. It might be some different loss. It is not necessarily telling from the picture, because if you have only two points you can’t make the example of “similar closer”. You can’t have an example involving three things. Your guess is as good as mine. Althought it is actually probably better because you actually read the paper [Laughs]

Interviewer: Is there anything you like about this diagram?

Participant: [Pause] It’s nice and colourful. I might like it more when I read the paper. I think there are very few diagrams of neural networks that speak for themselves, where you can immediately understand what is going on without any prior knowledge. I don’t know why they are doing it in 3D. I guess it is because these convolutional layers are 3 dimensional, but still, I don’t think you need to take that literally. Like what is the third, or second dimension for, for what it’s worth. You could just draw a line. I like the black silhouettes.

Interviewer: Anything else?

Participant: I do get that these two have the same parameters, but that’s how a siamese network works. It’s only telling you it’s shared. That’s it. I don’t know if this is shared or not, or if it is also a two stream FCN network.

Interviewer: The two streams, the bottom part?

Participant: Yeah, why is the top highlighted and the bottom not. To me it looks like it is the same, so why doesn’t it have a label. Is this also a two stream FCN, or are the two together the two stream FCN. Are those params actually shared? I think it is, but can only guess.

Participant: [C Pause] So the thing about these diagrams is they are nicely colourful. I suppose in this one it actually makes sense because they are showing a heatmap. I’m just missing the vocabulary here, I’m not sure what a disparity map is. I am not sure what guidenet is, I assume this is some kind of standard, like imagenet or alexnet or whatever. [Pause] I’m not sure what they mean by clouds. [Pause] I have a hard time telling what is the goal with what they are doing. Again, this is even further from what I am doing. I’d need to read the paper in detail to understand this diagram. Since they are using stereo input images, they are probably doing something in 3D, and the clouds are probably point clouds in this space. I think they are trying to get 3D out of stereo images. They’re having different cloud points and they fuse it, and by using this guided disparity map they filter it, probably to get a better representation. I don’t know if red and yellow or yellow and blue is distance. I don’t know, it looks like it is getting worse and getting filtered out. I don’t know. Since they have a car and the road, the task they are trying to do is something from autonomous driving. Again, I’m having a hard time understanding what is going on. I can see that over iterations things become clearer. It seems to be working for them, I don’t know. But it looks good enough from the pretrained guidenet, what it is spitting out. I guess an expert in 3D images could tell you more about it.

Interviewer: What things do you like about it? What don’t you like?

Participant: The colour is actually telling. Without colour it wouldn’t work. It is hard to say I don’t like anything, because I don’t understand it, right. I don’t like that I don’t understand it, I could say. But I don’t know how much of the diagram is really to blame. I can’t tell what I would’ve liked differently, because I don’t know what all these things mean, sortof. And I don’t know what I want to get out of this. Is it to understand how it works, is it to understand what their contribution is, is it what they are trying to do even [Laugh]. Based on that I would not like different things. But then again, it is probably unfair to expect from a diagram to explain to me what the goal is, when it would be relevant for someone working in that field. I can’t tell what it would be like for someone working in that field because I don’t have the knowledge. This diagram is definitely not suitable to communicate anything to anyone who is not deeply in the field, and probably read the paper. I guess I don’t like that, from where I am now.

Interviewer: Next one, D?

Participant: [D Pause] Right so this is some sort of natural language to SQL query, or some sort of query language. [Pause 20 secs] . That’s like some sort of sequence to sequence model, where they translate natural language into a formal language to execute a query when it’s done. As far as I understand, they have probably some gated mechanism to select when for this… Their vocabulary has two rules, select rule and select column, no three, and select table. Select table comes from some sort of schema encoder, I understand they just take the database and encode the schema somehow, just project it into a vector space, and also some sort of memory. Ah hold on, they also tag. Right so they also tag the question with a type, like if there is a match between the word and the column names it gets a different type. And then they have some sort of gated mechanism to decide if they select the column from memory or from the schema encoder, with memory being whatever they have in the query. But I don’t see the difference to be honest. Yeah. So, if this memory gate wasn’t here I would understand. I mean I understand the gate, there is probably some threshold, if it is above then select from one, and if below then select from the other, but why? I understand that memory, why doesn’t it have books? Because it is also, no that is only from column. But why, they are both. What I don’t like is that both book title and year are in both memory and in column. Why does it select book title from here and here. It might be something that whatever is in the memory is also in the column, but for that again I would probably need to read the paper. What is memory, what is this encoding. This I understand better than the CV examples, not necessarily in detail. If this memory is the only thing they changed from previous systems, they are not communicating well what their contribution is. But it looks like they did a bunch of stuff and just did not point out the difference between these two things. So yeah that’s natural language to formal language seq to seq approach, and you need some more things for when you select tables and columns from the database. Probably by some similarity between the vector here, that the seq2seq model is producing an encoded vector here, and the schema columns. I guess it plays a role whether they are encoding just by word or whether there is an overlap with the question, or not. I assume they are using biLSTM or a transformer to do both passes of encoding, that’s what the arrows both ways are here for, and while decoding they only go in one direction. I know that because I know how seq2seq works. I know that nowadays, when encoding the input, you usually take into account both directions, so that you have forward and backward information.

Interviewer: What do you like?

Participant: I appreciate that they have an instantiating example. This is more of a walkthrough of what happens to the data. So I like this diagram more. I don’t necessarily know what the encoder is doing internally. Internally it is probably some kind of whatever RNN GRU or something, it doesn’t really matter. This gives me an understanding of what the system is doing rather than what it looks like internally, what architectures. The architecture could be anything, while here I see what they are trying to achieve with it. They are not focusing on that, they are focusing on the process. I like that.

Interviewer: You’re liking the what is happening rather than the how is happening, is that fair?

Participant: Well, yeah because I would also argue that by looking at a diagram that describes the architecture you also do not understand how it is happening. You just see a bunch of vectors and things. I would definitely say yes, this is about what is happening.

Interviewer: And your preference is to understand what is happening?

Participant: Yes, because if I want to know how it is happening, the diagrams are not sufficient anyway. I need to look into the text. For example, I don’t know how this switch here works, right. And I would argue that it is probably really hard to sufficiently explain that in a diagram, so therefore if they try and fail it is wasted space or adding confusion. Or if they don’t, well in both cases you need to read up on it. And here they don’t even try. I mean there is some confusion to what is happening aswell. I guess what I don’t like is why is this memory exactly the same as part of the column, why is the memory only for columns and not for tables, why. Also what I don’t see from the diagram is how they trained it. This is at inference time, I assume. During training time it would look different. But then again usually diagrams don’t necessarily reflect that, right. If I take this and try to apply it, I know what will happen, but I don’t know how to get it that far. Maybe there is a different diagram that explains it. I can only guess they are using some beam search or greedy blah blah. The usual for RNN seq2seq models, but again that is more on the how you get this far.

Interviewer: So you’re interested in that aswell?

Participant: Yes, but I would probably get from text. Let’s say they are using beam search to train it. I need one diagram to understand beam search, but it doesn’t need to be in that paper. That’s when I go and google and try to understand beam search. And then they say we’re using beam search with k=5 and whatever and beam size this and that, I know what it means, because I’ve already seen one diagram for all the cases of beam search. Just as I don’t need need a diagram for seq2seq because I already saw a diagram that explains that, and that’s easier to save in your brain, because then you can instantiate “ah ok, seq2seq is some sort of sequence as input, and probably attention and blah blah blah and some sort of sequence as output”. I don’t need that explained in this diagram.

Interviewer: But it is explained in this diagram, is that what you’re saying?

Participant: No no no, it’s a good thing that they don’t go into details. It isn’t on a vector-number level how the seq2seq works. But then again someone from CV coming to this will probably have a hard time, just as I hard a hard time with all the CV diagrams, which is understandable because they are not using seq2seq and they are using an input as a sequence, not an image, right.

Interviewer: E

Participant: Wow what’s that [E Pause]. Ah I know this data set, it is something multimodel. Right, so I sortof can guess what they are doing because I know the dataset, and I just found the caption. So the input is a video and a transcript, so given this translate it into a recipe, a set of instructions to cook something that is described in the video. Now, they’re using BERT, which I know, to produce some sort of embedding for the transcript, and they combine the embedding with the video that is some pretrained resnet, I guess that is a standard for image embedding. Then they do some fusion between those two things. I’m not sure.

Interviewer: Within the context aware fusion module?

Participant: Yeah, here, probably it means frames as the video advances it goes to the right. And then they I think they get some sort of feature matrix out of that. And then at some point they form some sort of embedding of the fused input and the transcript for each frame, and they predict the procedure, where it starts and where it ends, and then they classify or caption that, by using those features and the transcript again. Again by using a sort of seq2seq thing but with features and some vectors. I appreciate this is a complex system, a complex AI system one could say, but what I don’t like is it is too much.

Interviewer: Too much… Content?

Participant: It is trying to explain too much, I guess. I think they are trying at the same time to describe the whole process while also highlighting the details. They don’t describe BERT, because that is not what they are trying to do. I assume this Process extraction and process capture module are their contribution. I guess it would be almost better to give a coarse overview of how the data flows. If that would be a black box, for the understanding of what is happening, it wouldn’t change anything. Both the modules. And then you would need an extra diagram to describe the module in detail. Or just refer to text, because. Yeah they are using some score feature embedding, but what does that mean. All those things don’t really mean a lot to me right now. Offset centre, offset length, average pool of selected frame features.

Interviewer: So this middling layer of abstraction is not quite working for you?

Participant: Not really, but I guess I know what all those things mean in a context, but here they are without context. “Position embedding” of what. I mean again there is probably some text somewhere which explains it, but without it. Also what is the feature matrix. I guess that comes out of the fusion, so how is it context aware. That’s it for my wisdom about this diagram.

Interviewer: What do you like about it?

Participant: I do like that you get the idea of what’s happening. It sort of does the job of explaining what is happening, up to the point where I understand it at a high level. If you ask me what is this architecture doing at a high level, I can tell you.

Interviewer: There’s a few different ways they are including dimensions of things. Self-attention 1 and n, but none between.

Participant: I assume they do have but just didn’t draw it out. I didn’t even realise.

Interviewer: So that’s fine?

Participant: I would be surprised if they used just 1 and n, they probably just forgot to put three points inbetween.

Interviewer: So an ellipsis would do?

Participant: Yeah. Once again, this is just BERT, because it wouldn’t need any of that. If they just write BERT that’s fine.

Interviewer: In terms of the visual encoding, here there is no problem with the self-attention layers, didn’t notice. Here are 10 or 20 circles. Is that significant, or could it be a different size?

Participant: I don’t understand what the sizes mean. Maybe 10 is because there are 10 classes coming out of resnet, I don’t know. Maybe their context awareness window is 10, so they look 10 frames ahead, or 5 ahead and 5 behind. I don’t know. It might be significant or it might be arbitrary, we’d probably need to check in the paper.

Interviewer: Does it bother you?

Participant: Not really, I try to. Since I don’t understand what it is doing I just skip over it. I guess if you. Now that you ask, I didn’t even notice. I said “I don’t understand this, I don’t even care”, right. Now that I need to guess, it sort of does bother me. But I wouldn’t start guessing on my own. Without being prompted to explain what any of this means I’d be like “I don’t know, I don’t care”. For me it doesn’t seem like it is important for the overall task of understanding what is going on, and for the diagrams that are following. Also, again if you start talking about this, why are these differently coloured, why is this blue why is this green, here I understand this is the span of the frames where the procedure is happening, but what does it mean here, why are those differently coloured. How does this relate to that. That’s what I mean, I understand what is going on but I don’t understand in detail. I think in order to describe how things are done, this diagram is not doing a good job.

Interviewer: Cool. Thank you. Did you like the module thing going on? The containers?

Participant: Well yeah in this case I think it sort of works, but also doesn’t because they have to replicate this thing here, the transcript. Why don’t they just use an arrow to put it here? Probably because it would introduce another line across the diagram. It sort of works, and is consistent with the exception of this thing. And again this is the same feature matrix, replicated again. I think instead of doing one crappy diagram they should’ve done three diagrams. One with the overall overview. Maybe even 4, I don’t understand what this context aware fusion module is doing, and it seems to be something worth mentioning.

Interviewer: Your final diagram, F.

Participant: [F Pause] Yeah that’s tix. That’s simple. Sort of.

Interviewer: Simple in a good way or a bad way?

Participant: In a good. That is abstract. I think what they are trying to do is score summary if it is good for a text, and they do that by projecting it into embedding space, running lstm over it, taking mean over time, merging it, whatever merging means. Concat or whatever. And then classifying or assigning a score in some way that is not defined. Yeah. [Pause] I do appreciate that it is simple. I guess I am having a bit of a hard time to understand the task they are trying to perform here, but probably by just the title of the paper I would know that, assuming the whole paper is about this architecture. [sees title]. I have to take that back, I don’t think the title will help me! Although they are trying to learn a scoring function. Trying to assign a score to a produced summary of a text. In this case, it is probably more interesting to know how they are training it and stuff like that, because the architecture is very simple. Like what is a good score. Here I think the training procedure is more interesting than the actual architecture. Yeah.

Interviewer: So, you mentioned this is quite simple. How do you feel about it aesthetically and in terms of the content here?

Participant: It looks good, it looks professional, I don’t like the ellipsis here for the merge. Also it is a bit redundant. “This is a fully connected layer, what are we using for the fully connected layer? A fully connected layer” . The duplicated labelling makes sense here and here, but not here.

Interviewer: Overall, which ones do you like and not like? Which is your favourite?

Participant: Don’t like A or B. Don’t like C. I do like D. I do like F, but I don’t think this is the main contribution of the paper, I think this is a byproduct. E I am not sure. I do appreciate they tried to describe the whole architecture in one diagram, but it would be better to split. I neither like nor don’t like. D is my favourite.

Interviewer: Thanks. The next part is a card sorting exercise. Please rank in terms of how important in your use of research paper diagrams. What would be good is getting top 2 or 3 right, and any you definitely don’t do.

Participant: [Pause] Identifying what it does I guess is identifying the purpose of the system. I don’t care about corpora or data types because I think that’s what the text is for. Memory aid yes. Right so the initial check I use the abstract for. The index to navigate is usually the introduction. I don’t think diagrams, at least as they are now, are not suitable for comparing systems, because as we’ve seen you can represent things vastly differently. Or completely different things quite similarly. [Pause]That’s interesting, what are the representational choices. I don’t think diagrams are capable of giving me a full understanding of how something works, just by themselves. [Pause] Yeah I don’t think the contribution of the paper is solely the architecture. [Pause] I mean, if they provide it is nice, but if a diagram highlights some specific arch features that’s nice and if I’m specifically looking for that I would also consult the text.

Interviewer: So the most important 3?

Participant: What it is trying to do, to have some picture in my head when I think about the architecture, and to be able to identify what they are using in terms of embeddings, projections, RNN, FC architecture, maybe what kind of loss. But again if training architecture.

Interviewer: Purpose. Memory. Rep choices. Layers and relations. And not at all: corpora. Initial check. Index. Comparing. Novelty. Altering. Happy? Great. So last part, do you use diagrams to reason as part of your creative process?

Participant: Yes. I sketch things. They don’t make any sense, but they do to me while I am sketching them.

Interviewer: When do you do that?

Participant: When I am thinking, not when I am actually doing it. Or when I am explaining something to other people, on a whiteboard. Usually using arrows to show where data goes or things go, conceptually. For myself, just to write down something, or to sketch down something really.

Interviewer: So your sketch will be of data moving?

Participant: Not necessarily. For me myself, the sketches do not make too much sense, I just start drawing lines to follow the thought process. Arguably they don’t really help, they don’t help after they are done, but while I’m thinking I just feel it helps to think about things by drawing them out.

Interviewer: And that’s pen and paper?

Participant: Yeah.

Interviewer: And you spoke also about communicating with others. That’s diagrammatic?

Participant: Yeah, and the good thing about that is doing it live. The picture at the end doesn’t need to make sense, but while you are looking at it while being drawn it might. You can use the temporality of it, right. You draw a line from here to there, and cross out something. In the end it doesn’t make sense, but the process of crossing something out underlines the thing you are trying to say.

Interviewer: Do you use diagrams in other ways during research?

Participant: Do memes count as diagrams? In presentations you use diagrams, but again you have progressive display. I think dynamic presentation. So you diagrammatically show some arrows and stuff. I think that helps, not necessarily to communicate in a static medium such as a paper, but when you are presenting to other people.

Interviewer: Is it different content to the paper?

Participant: Yeah, they are different. Noone thought about a dynamic paper yet. I think it would help actually, to have gifs in your papers. You can’t print it, but who says you need to print. [Laugh]

Interviewer: The next question is actually any tools that would help, so maybe this is related.

Participant: Yeah it would be an interesting concept to have. All those blog posts which explain architectures in detail, they all have animated pictures of how BERT goes and how conv layers how they work, and this attention. They are interactive up to a point, with mouseover changing colour based on how much attention and so on. And that’s helpful. It would be interesting to have that in a paper aswell. It seems to me it is a legacy of the past, and now we have computers as well as papers. A paper is a static format, right, but it doesn’t have to be. It also means that probably it would be more challenging to produce one. These blogs are made by people who devote a lot of time to do that. I don’t think architecture is the main contribution of any paper, but it can be something that is picked up and then carried over, for example with transformer paper. It was a good paper but it took a while for people to understand how to deploy it and how to use it in different ways. Might’ve been different if they had produced a more detailed. Or if they had released their code with the paper [Laugh]. In terms of understanding it, there is a very nice blog post which does a way better way of explaining how it works than the actual paper. So by using those gifs and 10000’s of images showing in detail how the architecture works.

Interviewer: Much earlier you spoke about no standards for these things. Do you think that would be beneficial or inhibiting creative process?

Participant: Hard to say really. I mean it worked for UML, there is a standard for how to produce a sequential diagram. I know how a sequential diagram works, if I look at one I understand immediately what is happening. But it is a very constrained case. UML data flow diagram, sequential diagram, you cannot explain arbitrary things with it. Research is not arbitrary, but I think it is not comparable to a sequence diagram in terms of complexity. You could have a standard, and constrain yourself to a standard, and it would cover maybe 90%, and then there will be edge cases which aren’t covered by the standard. I guess it needs somebody more principled to answer that question, whether you cover 90% of all the paper diagrams and the 10% you do whatever you want. Or you have a standard that nobody uses because it is not that easy to come up with a standard that covers all the complexity and is easy to use, and is better than the arbitrary stuff.

Interviewer: So in summary, standard diagrammatic language for AI systems?

Participant: I can’t say yes or no. [Pause] If you can identify a case that most of the things used, what the diagrams are used for, can be unified in a standard language, and that standard language is easy to use, and it is easy to understand and learn, then yes. Otherwise you know we have 13 competing standards, we introduce a 14th to unify all of them. Now we have 14 competing standards. I just don’t think it makes sense in an unconstrained case to try to catch them all.

Interviewer: Is there anything else?

Participant: No. You are most welcome.

# Participant 6 (P6)

[The interview started with some discussion on the participant's work. Note that the participant uses the word "graph" for all diagrams]

Interviewer: What is your experience level with CV/NLP?

Participant: Computer Vision I have 1 year experience. NLP I started 3 months ago. I’m a post doctoral researcher.

Interviewer: Your thesis was not on neural networks?

Participant: It was on fuel consumption, and I used neural networks. So I have a number of years experience in this.

Interviewer: Obviously this interview is about diagrams, so it would be good if you can describe how you use diagrams when communicating your research.

Participant: I just do a diagram which describes the system that I use. I assume you are asking about if I have a systematic way. I do not have a systematic way. I have one - it is called bond graph method, to describe energy flow in a mechanical system. But when I describe the model or architecture of the system, I just do it by intuition.

Interviewer: From nothing, you draw something?

Participant: There are some diagrams, like when you want to make a graph of a neural network you always pick neurons and you have arrows. You can produce a graph out of these, it is more intuitive.

Interviewer: Are you doing this based on specific examples you have seen from other authors, or is it based on convention?

Participant: No, I try to follow an example that I saw somewhere, in a book or an article. I don’t create new types of arrows or anything, I try to keep to convention.

Interviewer: Do you usually use diagrams in papers, for example?

Participant: Yes. I think it is necessary because it makes your idea clearer for the reader. Also, I don’t like having all text. I like to see a diagram which describes everything of the pipeline or the workflow.

Interviewer: So you are putting it in your paper for your readers’ benefit?

Participant: That’s one of the benefits. First of all, I always make a diagram during my research. I don’t see any other way. This is a kind of a plan. I don’t write a plan as a list, I write a plan as a diagram, where each part is connected. Which connections are important and so on. First I do a diagram for myself, and of course when I can correct this diagram, the final form of the diagram goes to this paper.

Interviewer: How do you decide what to include?

Participant: It is an iterative process. The first is always messy. And I do another one and another one and the final one is the final refined idea. It is iterative, that’s the method.

Interviewer: Is that on pen and paper?

Participant: Yes, it is much faster. I tried using some software but pen and paper is faster. I don’t want to waste time learning new software to make a graph that I will remove anyway. So yes, pen and paper is the best one.

Interviewer: And when you are writing up in a paper, which software do you use?

Participant: [Laugh] Last time I was using powerpoint because it was the easiest. I can imagine there is better software available for drawing diagrams but I didn’t find them and I didn’t want to spend time learning them. Just use powerpoint and that is all.

Interviewer: So you’ve described you like the pipeline and workflow. Does that include inputs and outputs, and is it all to the same level of abstraction?

Participant: It should be all at the same level of abstraction, right. The final graph should be the same level of abstraction but during the sketching part it can be different. Usually there is input and output. They are the features to start from. The left side is input, the right side is output, and what is happening inside is the question. First input and output, then what is the whole system or model. You need input and output because it is the base.

Interviewer: How do you use diagrams when you are reading a paper?

Participant: When I start reading a paper, I go through the whole paper, I see the figures, and where there is a diagram I focus on the diagram. Even before reading the abstract or conclusions. I go directly to the diagram, this is what I’m looking for, yes. What they do. The diagram should show you what they did in the paper, this is my opinion. Look at the graph, “Is it similar to what I’m looking for? Ok yes so I can read the abstract and conclusion, and maybe read the whole paper which will take me an hour”. The diagram is the main thing in the paper. From my perspective.

Interviewer: So you use it for initial screening? Do you use it while you are reading in any special way, or is it just part of the paper?

Participant: Yes. Sometimes it is not clear from the graph what is happening. Sometimes they include something from the graph, they say, what was it I saw yesterday, “metainformation features” or something like this. Something that was included in the model. We have no idea what this is, so we have to go through the text for this. But this graph should summarise the whole section of the method, what they do actually.

Interviewer: When you are navigating the paper, first you’ve found the diagram and there are some things you don’t understand about it, so you read the paper text. Do you then start at the beginning of the paper or use the diagram and say “oh I don’t know what this particular phrase means, so I will look this up in the paper”.

Participant: Yes, yes that’s the way. I look at the graph, then I look in the paper to find the features which are related to this particular feature on the graph. I don’t go from up to down in the paper, I try to find particular paragraphs which are related. When I read a CV paper for example, I am looking for the architecture they use, what was the input, what was the output, what are the methods, because you can use just CNN or usually include some novel idea which are the contribution. So you want to find in the paper what are these features, right. I don’t need to read the introduction. I don’t need to read the results. I just need to read what was the main idea and the graph should communicate the main idea.

Interviewer: Do you see any problems in consuming research more broadly in computer science?

Participant: Yes! Sometimes a paper is just too complicated to understand! Sometimes it is my fault, because I am inexperienced maybe. But in terms of communicating the idea there is always a problem if they do not have visual aids. I am more of a visual person, I like figures, I like graphs. Reading text is just too much for me. I like papers where the text is always supported by a figure. Sometimes people forget about it, or sometimes there is a limit on the pages, 8 or 12, so you just remove the figure because it takes up too much space. I think it shouldn’t be like this.

Interviewer: How would you feel about nanopublication, a relatively complete diagram being the entire paper?

Participant: What do you want to show the reader? Do you want to show only the idea? You can’t if you have no results.

Interviewer: Maybe a results table aswell, but something that would fit in one page of A4.

Participant: It could be some kind of invitation for reading the whole paper, because nowadays the research in computer science is very complicated, so if you publish something it is something very novel, which means it is something complicated compared to the previous works, so including all of this novelty of the idea in one graph… I don’t know if you can show it just in a graph without any text. I think it could be a good idea as a summary of the paper. If you could just summarise the whole paper in one A4 page with diagram and I don’t know results or conclusion. I believe if you present this architecture, if you present this idea, you have the results which support that. It could be a good idea. It could be much faster, to work with all the papers, to just look at one page of a diagram. This is what I do, but I have to go through all the pages to find the graph. If you just gave me one page this would be easier! [Laughs]

Interviewer: Examples. Tell me about it, what you like or don’t like about it.

Participant: [A Pause] What I like. [Pause] I don’t understand score with a dot in parentheses, I don’t know what it means, that’s the first question. [Pause] Ok dense layers, then we have a layer which are circles, I don’t know what that is. I don’t like the way it is presented, you have some text and some symbols and I don’t know what the symbols are. I don’t like it. Maybe when I read the whole paper.

Interviewer: Which symbols specifically?

Participant: Circles. Dense is a dense layer. 4 neurons. I don’t know. And then embedding layer, something similar with 5 circles. [Pause].

Interviewer: Is there anything you like?

Participant: The colours. If you can use colours, why not. The yellow and embedding layers, when you have yellow you know there is something from embedding layer. That’s all. It is not very clear.

Participant: [B Pause] Ok, convolutional blocks. [Pause].Ok that looks quite clear. [Pause] So with shared, I don’t know what that means, I assume it is explained in the text. [Pause] Yeah I think that looks good, because I think it is the convention that you represent these convolutional blocks in 3D like blocks. So it is quite clear. [Pause]. I don’t understand the contrastive loss. Maybe that should be a bit clearer, I don’t understand the blue dots. Is it output. That’s what I don’t like.

Interviewer: The numbers on the blocks?

Participant: This is the depth of the conv block, but we don’t have the number of pixels. I mean this is the third dimension. There are another two dimensions I don’t know. Maybe they should be included here.

Interviewer: Is this the type of information you’d like to have in the diagram, including these depths?

Participant: Without the depth it is just meaningless. It would just be a block, and I don’t know the dimensions of the block. You can see they are different 64, 256. Maybe they should be 3 dimensions. [Pause]. Yeah it should be.

Interviewer: So overall?

Participant: Better than the previous but still there are a few improvements.

Interviewer: It sounds like there are conventions which are helping you to understand?

Participant: Yes, yes. When I see the block with a third dimension 64, 128, I assume that this is a CNN immediately, because that is the way it is presented. Sometimes in 3D, sometimes in 2D it is not so obvious.

Interviewer: The shared label, would you want that to be more descriptive there, or is it good enough as an index to the text?

Participant: I think if they include that it is shared using some arrows, it would become less clear. I think just shared and described in the text is enough.

Interviewer: The use of colour?

Participant: [Pause] They are using, but I don’t know what the colour represents. I assume that they are different layers, different blocks with the same dimensions but different colours I don’t know why. Maybe a legend would be a good idea. And it doesn’t say in the caption, right. No. So they use colours but they don’t say why.

Interviewer: C?

Participant: [C Pause] I like that you can follow the workflow and see the result of each step. You have the base image, then the pretrained guidenet, and the result is this image. And you have some sort of primary filtering, and you see the result. Secondary filtering and you see the result. That’s a cool idea. [Pause]. I think it is quite clear. I don’t know the symbols like Hg. But I assume they are described in the text. [Pause] Iteration 0 and 1, why is this? [Pause] I don’t see the final result, what we want to achieve. I believe the output is the image after the second iteration, so the input and output is not defined here. Stereo input, ok, but the clouds, are they another input or something additional, I don’t know. And the output, there is no output, some something after the second iteration is the output. It is not clear, so that’s what I don’t like.

Interviewer: Anything else?

Participant: Pretrained and finetuned guidenet have the same symbol.

Interviewer: Have you seen a diagram like this before?

Participant: I don’t remember to be honest. I like the idea that you see the result of each step on the image. That’s quite cool.

Participant: [D Pause] Ok so again, where is the input where is the output. At first glance you have no idea. I assume the bottom is the input, but no idea what the output is. Encoder. Decoder. [Pause] Yeah this is the kind of graph which I need to read the text to understand. [Pause] A lot happening here, and the result, I assume they use memory to have book title. Yeah for me it is not very clear. Maybe if I would use this kind of architecture it would be more clear. Just looking at the graph it is hard to understand.

Interviewer: Anything you like?

Participant: [Pause] Again it is all on an example, I think this is good. It is not just the overall idea, but it has an example. Similar to the previous. This is good. But the way it is presented could be a bit better.

Interviewer: And the colours?

Participant: They are good because you have a legend. At least it is explained why this colour is used here.

Participant: [E Pause] Main structure of our model. It looks quite clear, to be honest. [Pause] We have input and output, very clearly defined. [Pause] Yeah I like the structure of this diagram, I think it quite clear. When I read the explanation in the text, I think I would understand it very well after reading the text too. [Pause] The colours are also used, I think in a good way, for different parts of the model. [Pause] Do you want me to say more?

Interviewer: I’d like to ask some specific questions. Do you think that self attention layer 1 and n, does it bother you they’ve missed the ones in the middle?

Participant: Not really. I assume this is the kind of architecture they use. There should be something in the middle.

Interviewer: There are n layers, they’ve drawn two of them.

Participant: Ah ok [Laughs] Now I see they should show that, or at least three dots.

Interviewer: Is that annoying for you is what I’m trying to understand?

Participant: Yes, because it is inconsistent. There is a convention, if you have many layers you use dots to show there are more. You mentioned that, I didn’t see it. I missed that. So that means it is not good, because I missed that. It should be included here, and it annoys me yes.

Interviewer: Another question. Here there are two columns of 10 circles. Do you think that 10 is an important number, or could they have used fewer or more? I’m interested in your thoughts on whether being precise about this is annoying?

Participant: They could include the number. The graph is ok, but they could write the number of the layers or neurons or whatever it is. Maybe this is valuable information, I don’t want to count them myself. Similar here for the features, what is the dimension of the feature matrix? I don’t want to count.

Interviewer: Do you think it likely, I appreciate you haven’t read the paper, that 6 and 13, do you think that maps directly to what it would be in the code?

Participant: Yes, because it looks like the dimensions are fixed. I think in the code it will be a 6 by 13 matrix, and that should be written down here.

Interviewer: So when something is written in a precise way, with like before 4 circles, do you think that means there are 4 of them?

Participant: Yes! I mean this is computer science not literature. If you have four, it means you have four. [Laughs]. It’s clear, right. If you have n or m, it means that it varies. But if you have four it means you have four, that is clear for me. If you show on a graph that you have five layers you cannot use seven layers [Laughs], you cannot do that. Four equals four, that is clear.

Interviewer: So the amount of content in this diagram is quite a lot. Is it managable?

Participant: Yes there is a quite a lot. The model is complicated, but you can follow the arrows, and you can follow all the parts of the graph, and I think at the same time you can follow the text and I’m sure that’s clear. [Pause]. It is manageable, yes.

Interviewer: It also has this macroscopic module structure over the top, do you like that?

Participant: Yes, because this is each step in the flow. If you can divide your module into such steps then I think that is a good idea. If it is possible of course. Input, preprocessing, ML algorithm, I don’t know some feature extraction, then something else then output. If you can use modules to represent it, it is clearer for the reader.

Interviewer: Now F.

Participant: [F Pause] That’s the easy one. It is quite clear. It misses a lot of details, but as an overview of the model it is very clear. [Pause]. Input output, layers. As I said, there are not many details here, what is the architecture of this LSTM, what is the projection layer, what are the dimensions, how they merge the results from previous layers. What are the layers. And softmax. And output is score but how many labels we have. This is a good overview but very hard to say what the model does. It is classification right, but into what. A lot of information is missing.

Interviewer: Do you think this has enough information in it for you to be able to look it up in a paper?

Participant: I think we need another, more detailed graph to represent the architecture of the model. This is just an overview. Maybe it is a very short paper this.

Interviewer: Is there anything you like about it?

Participant: Yeah, if you just want to present the main idea of what the system is doing, this is very clear, that’s fine. Maybe this is a good idea to present in the paper before the whole model like presented in the previous graph. Then you include all the dimensions and feature extractions and methods. This is just to give the idea of what is going on here, and if you are interested in more you can go to the more complicated diagram of the architecture.

Interviewer: So these labels at the top of the diagram, do they help?

Participant: Yes. Maybe they could be a bit better defined. Merging layer is just the merge, or merge plus mean over time, I don’t know. FC layer, might just be FC layer. I don’t know, because it is not so clear.

Interviewer: We’ll move on to the tasks now. If you could order them in terms of important and not important. Most important in your use of diagrams. It would be good if you can sort the top 2 or 3, and let me know about any you don’t do.

Participant: [Long pause while reading and sorting] I have two most important ones, and three which I don’t use at all. The most important are initial check, and understanding how the system works. This is the same as I said at the beginning. Least important is identifying opportunities to alter the architecture, identifying what the author thinks, and memory aid.

Interviewer: Anything else you’d like to say about how you use diagrams or what tasks you are performing?

Participant: This index to navigate the paper, is also useful but not crucial. Purpose of the system, you can see from input and output. And you can compare two systems but comparing two systems is difficult because everyone uses different graphs, so when you compare two graphs you have to understand the whole system anyway. Comparing two graphs is just useless because they are not uniform. They are a different level of abstraction, they have different layouts, different ways of representing the system. You have to understand the system to be able to compare it.

Interviewer: The last part is some broader questions, you’ve already covered quite a lot of this already but I want to give you the chance to say anything else. Do you use diagrams to reason as part of your creative process?

Participant: Yes. Absolutely.

Interviewer: Is there anything else in that area?

Participant: Yes. For me it is hard to imagine doing this in a different way. I mean, how to develop a new idea without doing a graph of your idea. I just don’t know how to do it any other way. Writing a text is just, you need to make a graph.

Interviewer: Do you use diagrams in other way during your research more broadly? Like drawing on a whiteboard to communicate?

Participant: Usually I use just pen and paper, because I don’t have a whiteboard, but if I had I would use. There was a time I had a whiteboard, and I was using it quite frequently, because it is a bigger space for representing this idea.

Interviewer: Was that for your own reasoning or to communicate with others?

Participant: Both. First of all my own reasoning, and then to other people, but less frequently. Mostly for myself.

Interviewer: When you’re communicating, would you want to draw while you are talking, or to have the diagram finished and use the diagram to talk to?

Participant: It is better to do the diagram while talking, but the quality of the diagram is then not good enough sometimes. In terms of aesthetics, it is better to have a prepared diagram, but it is easier to follow your thoughts while you are drawing the diagram. Usually I don’t have a prepared diagram, so I always draw this [Laughs].

Interviewer: Makes sense. Are there any tools that you think might help you to make or consume diagrams?

Participant: I’d really like to find a software which is efficient in doing diagrams, but I didn’t find this yet.

Interviewer: What do you mean by efficient, exactly?

Participant: Easy to use, easy to learn, and it doesn’t constrain. You can always use paint, but it is painful to draw with paint. Sometimes when you use software for drawing graphs, it turns out you cannot do something because there is not this kind of arrow or this kind of something, and you spend more time following the way the software is written than using your creativity to do the diagrams. That’s the main problem, I don’t know how to solve it, it is not easy to write this kind of software.

Interviewer: Would you want a standard language for diagrams, or is that overly restrictive?

Participant: [Pause] Just for the sketch of an idea, I think it is not necessary, but the standard way of presenting a diagram in a research paper is a very good idea, especially in AI. It should be something like this.

Interviewer: Why especially in AI?

Participant: Because it is quite novel, and for example in mech eng you have methods to show the flow of energy and it was developed 20 years ago and people use it. In AI every paper is presenting a system, usually, and they present it using a diagram, and everyone is using a different way of presenting it. It would be nice if it was a uniform way of presenting it.

Interviewer: Do you think there is a danger that new things are coming out so quickly, new approaches, that the diagrams would not be able to keep up. Like UML can’t keep up with new things.

Participant: Yes, there is a risk, but there are some features which are constant. Input, output, you always have it. What is in the middle, that is the question, that is not constrained. But we can stick to some uniform symbols or legend. If you do something new, it becomes even clearer what is the contribution. This is standard. These are standard features we have been using so far, and these colourful new symbols are our contribution. And if this idea is cited enough times in the literature then you include these new symbols in the uniform code or convention.

Interviewer: Anything else you’d like to say about diagrams before we finish?

Participant: I like diagrams. [Laughs]

# Participant 7 (P7)

Interviewer: How would you describe you describe your experience level with CV and NLP?

Participant: Computer vision I’m just reading papers, but I have not really tried out anything tangible. NLP I work with for tasks like NER and POS tagging, question answering, natural language inference. A lot, I would say extensive NLP. I’m in the second year of my PhD in natural language processing.

Interviewer: How do you use diagrams when communicating your research?

Participant: Most of the research we do, we deploy some sort of computational model, so we use diagrams to express those models quite easily. Rather than reading a bunch of text, it is much easier to, for example in a neural network model, to express it as a diagram. It is much better. And in some cases I use a diagram to show the whole workflow of my approach, and for that I employ a diagram as well.

Interviewer: Why do you use a diagram, as opposed to text?

Participant: Because in the diagram you can express directionality and you can show consistent things over the whole approach. You can have a holistic view in a diagram, that is quite hard to do in text, there is a bunch of stuff you have to go through to understand. For a new reader, providing a diagram that gives you most of the picture.

Interviewer: How do you decide what to include in your diagrams?

Participant: It is a bit tricky I would say, because we are inspired by other approaches. We don’t know whether to take everything in from other approaches, or to do a part of it and do it. Most of the time I take the computational models not in deep detail but in a small box or something like that. And giving that a name, hopefully the readers will pick it up in the text in detail about that.

Interviewer: So you’d explain BERT with a little box saying BERT? And you try to do that in general when you build on someone else’s work?

Participant: Yes.

Interviewer: So how do you decide how to represent what you want to represent?

Participant: Mostly we identify the main blocks, building blocks, of my approaches. Most of my research is considered towards building a methodological approach, so I will work towards identifying the important building blocks, but it isn’t anything streamlined, it is more of an intuition. Then I try to express those building blocks in a diagram, yes.

Interviewer: Do you tend to draw block diagrams, that are like input, output? Are they literally squares around words?

Participant: Yes, squares around words, and sometimes those squares might have numbers to say these are the hyperparameters we used or this is the dimension of this, and that’s it. So it is mostly squares.

Interviewer: You don’t use the iconic representation of nodes?

Participant: When it comes to graphs, we use iconic representations. When it comes to neural network diagrams, it is more arbitrary. Or I mostly go through NeurIPs of ACL papers, and see which diagram I understand better, and take the same process in my diagram, and put my vision into it.

Interviewer: And what sorts of problems do you encounter when making diagrams?

Participant: Lack of consistency.

Interviewer: When you’re making it?

Participant: Yes, because we don’t have a language for that, and how we express things is not really common. It’s not there. So it is very hard. I might intend something in my mind, and my diagram might be interpreted in some other way by another person, so we can have a mismatch there. I think this lack of language for diagrams is so bad, even at a high level there is nothing the same at all. Every time I use an approach, I use different colours, different blocks, different ways, because it is arbitrary for me when I do it. When I look across different papers I have written so far, the diagrams are very different, each one. When you look at the diagrams they are different, but the approach, the way it has been composed is the same. The colours the diagrams the diagrams, it changes.

Interviewer: Ok so it’s the same or similar content, but the visual encoding is different, because it is..

Participant: It is whatever I feel. Whatever I feel is tried I will do it.

Interviewer: So you have a supervisor, does your supervisor provide feedback? Do you show your friends?

Participant: I do show it to my supervisor, but he was pushing me for using a diagrammatic language which has not matured yet, so I am not using it yet. So I do show it to my friends and give me some comments regarding, they may give some more subjective comments. I evaluate their subjective comments and I make my subjective decision to put it there.

Interviewer: So when you are drawing, you mentioned it might be a system or a workflow diagram. Does it encompass the inputs and outputs aswell?

Participant: More or less it has that, every diagram I have drawn so far or seen so far in a paper has an input and output. That’s what it is geared towards. We are computer scientists, we are interested in the input and output.

Interviewer: What tools do you use to make these?

Participant: Draw.io. Exclusively.

Interviewer: Why?

Participant: Ease of use, pretty straightforward. It is mostly ease of use and there are lots of pre-templated things that you can use. So I use that.

Interviewer: If you had a diagram that you’re drawing, and there is some part that you want to describe in detail because it is your contribution, how do you decide what level of granularity to go down to, because you’re not going to code level.

Participant: So, I have an internal process in my mind, when I come up with a method I like to have this in a diagrammatic way. I think in drawing stuff. That’s how I do it, it isn’t writing stuff. Even when I come up with something, I draw it, and it varies a lot. When I come up with an approach or something like that. Then I go into granularity where my contribution is. So for example if I have a GNN approach, and I’m making a small change, I make sure that I go into detail there, to make sure my contribution is shown. So the granularity will be based on the contribution level, so it varies.

Interviewer: You were describing there using diagrams for your own creative process, is that pen and paper?

Participant: Yes, pen and paper diagrams. I usually draw on paper and then put into draw.io.

Interviewer: Ah, so this is before you’re even writing the code?

Participant: Yes. But sometimes I draw the diagram, and the code changes a lot. The initial version is different. That’s what I do mostly, because that gives me a lot of clarity.

Interviewer: How do you use diagrams when consuming research?

Participant: For example, if it is an analysis paper or a results paper go for the results section first. If it is an NN paper, the approach paper, because that will give a cross cut of where the contribution lies. After reading the diagram, I go and read the rest of the paper. Reading the paper starts with the diagram, for me.

Interviewer: Do you read the abstract?

Participant: Yeah I skim the abstract for keywords, and if those keywords match my interest I read the abstract more completely. Then I go for the diagrams first.

Interviewer: Then what?

Participant: I read the diagram and see where their contribution lies. Most probably I do understand their contribution from the diagram, but not fully, so to fill in the gaps I read the paper.

Interviewer: From the beginning?

Participant: I go to the method. It is variable, but if I had a block step-by-step thing: Abstract cursory reading, abstract depth reading, diagrams, a bit of introduction, a bit of methodology to see, and if I see it is interesting then I go for results as well.

Interviewer: Do you interact with the diagram at all, in the paper, do you draw on it?

Participant: No.

Interviewer: What problems are there in consuming research more broadly?

Participant: That’s very broad. I can’t answer for every area, but for my research area, for my PhD, it is I would say lack of clustering. For example people are in different conferences, different journals, people are working on different stuff, and they are publishing works. If you look at a birds-eye view, you could see patterns, if you cluster it together. And this can benefit from each other but it’s all very isolated and they do their own thing and never converge on anything tangible. So that means that finding, going round and finding research is very hard. Because you might see something and people see a gap in a research field, and in a different domain a difference place has done it. It’s very vague, and there is a lack of unity there.

Interviewer: Now the fun part, we have six diagrams. Tell me about it, what you like and don’t like.

Participant: [A Pause] Ok. So, since I come from a more similar field and I understand this diagram well, I like this diagram. I can understand the dense layer, the hidden layer, the scoring layer, the softmax, the sigmoid and how it back propagates in every layer. The author does clearly say to me. If I read the paper and I see this diagram, I can clearly say where the contribution will lie in, and what sort of contribution the author is going to give to us, there. So then I can study in the paper in much more depth. For example this dense layer, the same shapes are throughout, but this and this doesn’t have the same properties. This is like a layer, this is like a kind of layer, but it is different and same shape. And show h1 and hk are hidden layers, this should be openly aware for an ML or deep learning guy, but for a linguistic person or someone else they might not be aware of what the numbers are here. And so that will not be clear. They are using genre aware attention, so one of the things I feel is that in these approaches it is not only about the approach, it is also about the hyperparameters, and that is not there. The dimensions are not there. Because when I try to write my own deep learning code I tend to see the code in numbers, in matrix numbers. Because it is mostly matrix multiplication, or transpose, it’s always numbers and that is always easier to understand the code, that isn’t written here. There should be loss at the end, which will depend and they’ve not mentioned it here. So yeah.

Interviewer: So there’s quite a lot of stuff there. Would you want different shaped boxes for sigmoid and dense?

Participant: If you think, those are different things, conceptually different things, right. So yeah you train.This will change. Sigmoid is just a sigmoid transformation. But you train the model. So the diagram should express that, surely. For me I know what is sigmoid and dense but when I started looking at deep learning diagrams I was really confused about this. “Then sigmoid should also be trained, but it is not, it is just a sigmoid function”. So I try to cluster things so it is easy to understand. This gives a very wrong impression, I would say. Here, these dense layers, they change, this is where you would change the numbers. The softmax and sigmoid will not change. There should also be what parameters they are tuning, the parameters should be here if I am right. At least saying these are the parameters we are tuning, and these are functions, you know, which will not be back propagated and changed. The commonality is not here.

Interviewer: You’d like that to belong in the diagram, not the text?

Participant: In the text as well, but in the diagram also it would be much better I’d say, because then you know for a fact when the author, this approach is trying to tune this dense layer, this is just how to get to that point.

Interviewer: The embedding layer?

Participant: That’s vague. What does that mean. Dense layer is very canonical, but the embedding layer could be anything, could be BERT, could be anything. It doesn’t give me anything there.

Interviewer: So you’re after the specific way it has been embedded?

Participant: Yes.

Interviewer: Do you like the labelling on the sides?

Participant: Yes, that really helps. But still, focus on inputs, there is a lack of consistency on input layer. Ideally it should be input layer, dense layer, merge layer, classification layer. When I look at this lack of consistency there, I will need to look at the paper to clarify that. Ideally I would like to get at least 70% of the approach by looking at the diagram.

Interviewer: Does the inconsistency of the labels here annoy you?

Participant: Yeah. [Pause].

Interviewer: Have you seen diagrams like this one before?

Participant: Yes, lots. Lots and lots. But with different colour coding. Different shapes, different colours, different dimensions, different everything. But the same family I would say.

Interviewer: So in the hidden layer with h1 etc, it has 4 circles in a box.

Participant: Doesn’t mean anything. They came up with their own colour coding, that’s it. And the number, it cannot be four dimensions, it must be much more than that. It is either misleading or just plain their own internal reasoning to put it like that.

Interviewer: And the embedding has five, rather than 4, do you think that is significant?

Participant: I would hope the authors put it in a way that has some internal reasoning to that. Maybe if I read the paper I would understand the reasons, but I don’t.

Interviewer: So you’d go to the paper to understand?

Participant: Yeah, I would see 4 and 5, there must be a reason, this layer should be a bigger one. Or maybe the hidden layer has some other transformation layer, then maybe that is why they put that. I’d go in and check it out in the paper.

Interviewer: Next one, B.

Participant: [B Pause] Ok. So, I really like this one. I would say I even like it better than the other one, because I can see the dimensions and I can think it through. The problem is for example I assume they are using a CNN, and they backpropagate using siamese network. Since I knew what a CNN is, and how a siamese network works, I can understand it. At least abstractly. This more fine-grained, I would like more to incorporate with this kind of level of abstraction. Ok, this kind of diagram, they use CNN, conv1, conv2, I assume it is CNN, but for a new person when they come in I would be confused by this one. If I had CNN, which is canonical, there. If this diagram was accompanied by a more abstract one, that would be much more helpful.

Interviewer: So a more abstract version that says CNN? Not showing each layer?

Participant: No, these layers as well, but also saying a very simple “input here, CNN here, then finetune siamese network like this”. Then I have go a level up to think for myself, which is easier for me and that is why I like this one much more, since I know the field, I can abstract it. But for a new person maybe this is difficult. Yeah.

Interviewer: So you like the hyperparams?

Participant: I really love it.

Interviewer: This shared label?

Participant: Yes, that comes from the siamese network. If you know siamese network, this makes sense. Otherwise, not. I think you should also have that abstractness represented as well. Maybe you can put a box here with CNN, and have these internal things here with an overlay something akin to this one, and overlay it.

Interviewer: What do you think the colours mean?

Participant: I don’t know. Just random for me. The paper might have something. Since they are pictures maybe they are the same thing boiled down to black and white pictures, but I don’t see a reason there.

Interviewer: What’s your opinion on the contrastive loss representation?

Participant: Umm, I don’t understand. So this is good, because this was not something in the previous. They explicitly said loss. The content is in the paper. The loss is contrastive loss, and is an important part of that process, so that is good. They say what space they do it in, what embedding space, but I’m not really sure why they put two blue dots and a line there.

Interviewer: Overall?

Participant: I like it, I really like it.

Participant: [C Pause] Ok so I am indifferent to this. One thing I like is that the diagram blocks have some meaning to it, unlike the colours they have chosen in other approaches. They show ok whatever this do and what they look like. The colours and the blocks have meaning and that is really great here. But at the same time they have these two inverted triangles, whatever they are I don’t know why. The guidenet part. And the projections, what will the dimensions of this image look like in matrix form, and what will it look like here. How the approach works is not there I would say. Maybe a computer vision expert would understand this, as there might be some coding behind that. Because I’m not a computer vision person I’ve not seen the code yet. Still this is out of place, I would say. And there is no direction to the diagram, in terms of where you start and end. Maybe there is, but it is not clear for me. Because the iterations.

Interviewer: This uses one example through it, do you like that?

Participant: In my diagram no. If I was making it, I don’t like a running example. A running example, maybe it is nice when it is used. Yeah, I think a running example is good.

Interviewer: This is quite dominated by an example. A lot of the content is about that example. In this particular instance do you like that?

Participant: In this particular instance, [Pause] yes. Because it is about images, and how that works is fairly clearly represented. So if I in a paper had this running example introduced from the beginning, and then I can take it to a diagram and run it through a diagram, I can understand it much more easily. I think that is what the author was going for. Now I know what they’ve done, I can go back and find a running example, so I would say yes.

Interviewer: Anything else? Overall do you like it?

Participant: I’m indifferent.

Participant: [D Pause] So. I like the fact that embedding encoding this legend is useful to understand the commonality. There particular contribution is in this level, and for that they’ve come up with their own way of expressing there. It’s new to me, so I have to read the caption to understand. For me I don’t see any other way of doing it as well. I like the running example again here, and they’ve said a lot about what they are doing, I like that part. So overall I like it, even though it is particular to their particular contribution, I like it.

Interviewer: Is there anything you don’t like?

Participant: There are these particular rules, right. Maybe I would have to read the paper to understand that. But in a real scenario, you should be able to abstract it. To up the abstraction to give a different name here, that might be easily identifiable by anyone. But then again if your contribution is on these things you have to put it there, there is no other way around. So I don’t like it. This part is a bit disjointed, even though it is a clever way of saying memory on, memory off, but still I don’t understand why. But I’d read the paper to find it out. Overall I like it.

Participant: [E Pause] So yeah. I would say it is too much in the diagram for too little information. Why do they choose these different colours, I don’t understand this part. This lack of direction here. I like how they have put examples here, so at least I have a start and an end, but there is a lack of directionality in the diagram. They use pretrained models, but for here they have a feature matrix which is used as an output, but for here it is simple, so there is a lack of consistency between this one and these ones here. The self attention layer and the feature matrix. Why these zeros here in the context aware fusion model? Maybe if I read the paper I would understand, but the diagram I don’t. So there are a lot of squares, a lot of colours, but why, I don’t understand. I’m not a big fan of it.

Interviewer: So the overall structure has these modules.

Participant: Which are good, but since you’re going into so much detail it would be worthwhile to know what parameters they are using, and why they put these circles here because it doesn’t mean anything to me. It is random, in a way. I feel a lot of randomness here, no clear choices I can identify. Maybe if I had the paper I could identify the choices they had made in their diagram.

Interviewer: What do you mean by it feels a bit random?

Participant: There are circles here, and green colours there, are these coming from here, these blue ones where has that come from, there should be some reason to that one. There must be some reason to that, I’m sure. And this colour and this colour match, there must be some reason to that. So my intuition is that the author did not write the diagram to stand on it’s own. More like you read the paper and then the diagram will help you to understand the paper much easier. This diagram does not stand on its own.

Interviewer: Some of the arbitrary things?

Participant: Yeah these arrows, these shapes, the number of arrows, why three arrows. And these circles, this slash green thing and this dot and this comma. There are two green colours here, are these pretrained things are your contribution is here so.

Interviewer: So overall?

Participant: Don’t like it.

Participant: [F Pause] Again so I’m indifferent to this one. It expresses what it intends, so you can see everything explained in the paper in this approach. I see enough to understand. But again, what are the dimensions here. It is too abstract, and does not give concrete numbers here, which I would like to have seen. I am indifferent to this one mostly. The projection layer, the mean over time I would assume is a function. These are the trained ones, so they used the same boxes here, you have to train this one as well.

Interviewer: The labels?

Participant: It helps, it has some consistency. It conveys what it is supposed to, so I understand. It’s not like the one in E there but again it is a more complicated model, I can understand it perfectly enough.

Interviewer: So there are a few overall questions. How important are aesthetics to you in diagrams?

Participant: Umm, a colourful diagram is always good I guess, yes. But I’m not pushing for that but it would be nice if it was colourful and had nice aesthetics. I think the content is much more important than aesthetics.

Interviewer: Which one was your favourite?

Participant: Umm, B.

Interviewer: Cool. Are there any types of diagrams you’ve seen that you think are not useful?

Participant: Nothing comes to mind. All the diagrams I have seen so far convey some kind of message so it is fine, it is always useful.

Interviewer: Do you like to see diagrams in papers?

Participant: Yes, I love to see diagrams in papers. More the diagrams I would say.

Interviewer: More the diagrams than the text of the paper?

Participant: Yes.

Interviewer: Wow. So when thinking about the examples, you mentioned liking the persistent example, is it sufficient to have at the beginning and the end, or do you want the middle as well?

Participant: Certainly beginning and end, but if it makes sense in the middle as well.

Interviewer: Anything else?

Participant: No. It’s different, so it has also given me a perspective on how I should approach diagrams now [Laugh].

Interviewer: An unintended benefit. Final exercise, a list of tasks. These are things you might be doing with diagrams. Can you rank in terms of how important, top 2 or 3 in the correct order, and anything you do not do. The ones in the middle don’t worry about. And if something you want isn’t there, we can write a new one.

Participant: [Pause] Memory aid I am not interested in using, so that one is fine. [Long pause sorting cards] These I definitely do, these I do sometimes but not that much, and memory aid is the one I don’t do. Top ones are initial check, understanding how, system novelty, representational choices, comparing, specific features, opportunities to alter, relations between componenets, external dependencies.

Interviewer: Did this list feel complete?

Participant: Yes, most definitely.

Interviewer: Do you find you can get what you need, in general, from diagrams in papers?

Participant: Overall yes, but I would like more information, it is very misleading. But I take a pragmatic view when I’m viewing diagrams, and I don’t try to identify them with commands, I go with a more inquisitive eye and get on with it fine. But I would like more information.

Interviewer: Ok so it’s almost like you’re using the diagram as a way to ask questions of the paper. Does it anchor your [interruption]

Participant: Yes for me yes.

Interviewer: So the last part is some extra bits, we’ve already spoken about much of this. Do you use diagrams to reason as part of your creative process?

Participant: 100% every time.

Interviewer: Anything in addition to what you mentioned earlier?

Participant: Not only on my current problems, but keeping track of my work as well. I take “I will do this part, I will do this part” and can add what are the problems on this part. The diagram is my thinking process. But it is random. The way I compose the diagram is random.

Interviewer: So are you quite visual?

Participant: Yes, but I’m not a visually gifted person.

Interviewer: Rather than audio?

Participant: Yes. I think this comes from back home, during exams all the time I try to draw stuff to make it make sense.

Interviewer: Do you use diagrams in any other way?

Participant: To explain things, particularly to my supervisor. To explain what I am doing, it makes it much easier to explain. I can start from input and output. So I have this direction in mind, and I can explain to him this is where I start, and then come here and come here and this part has these problems here and these parts have an output so we have directionality of thought as well.

Interviewer: Is that on a whiteboard or on paper?

Participant: Mostly on whiteboard, sometimes I put in draw.io and print it out. Sometimes on paper.

Interviewer: Can you think of any tools that would help you when making or consuming diagrams?

Participant: I am pretty happy with draw.io, but it more generic and not geared towards research or machine learning research. I would like to see some of those components integrated into draw.io. So they have flowcharts they have all the boxes, vision box, turning box. For NNs if I could pick something out and drop it there, that would be amazing.

Interviewer: So you mentioned earlier about there not being a standard way of doing stuff. Would you want to have something like UML for AI diagrams?

Participant: I’m not sure if… I would love it to be there, something like UML, but if I would take to it like a duck takes to water, I am not yet entirely sure, as it might be too much of an effort for me to understand this. Maybe a simpler version of that would be good.

Interviewer: So ease of use is important?

Participant: Yes, because I don’t want to spend my time, a week or two weeks, to understand the diagram itself. So ease of use, but to have consistency.

Interviewer: Is the inconsistency causing problems for you?

Participant: Definitely, because it throws my train of thinking. If there is a good NN diagram then it streamlines my thoughts. Sometimes if there is randomness to it, I fail to see the patterns because of the high amount of randomness in the literature. And also in my thoughts, which are random in themselves, but when I try to put them in a diagram there is a lack of commonality there. Too scattered.

Interviewer: Is there anything else you’d like to say about diagrams before we finish?

Participant: I love diagrams. But I mean it. [Laugh] That’s it.

# Participant 8 (P8)

[This interview had connectivity issues, particularly during the first parts. During the recording this was noted and aspects were asked to be repeated. This transcript attempts to record the content and intent associated with the interview].

Interviewer: Tell me about yourself and your experience with CV and NLP?

Participant: You know my name is [redacted for anonymity], I’m a researcher in AI, specifically NLP. My experience with CV is quite limited, I had one or two courses with CV trying to achieve some image detection using NNs. Relating to CV my experience is limited to object detection and identification. Speaking of CNNs, RNNs etc I have been working with them for a while. Using textual embeddings and also for tasks like semantic processing, some language models. So this is basically my experience with neural networks. [unclear]

Interviewer: It sounds like you have deep NLP and some CV knowledge, is that right?

Participant: Yes. Lots of NLP and some CV.

Interviewer: Perfect. So could you describe how you use diagrams when communicating your research?

Participant: In a general sense or in a NN?

Interviewer: I’m interested in a general sense, but NNs and their system architectures are of particular interest. Whatever diagrams you do are interesting.

Participant: Ok. In the general sense I use them when I want to express something that would be better expressed or [unintelligible].The idea with a diagram is to identify the main point [unintelligible]... high level representation of that particular technique … idea in a high level … pipeline or… [unintelligible].

[Interruption for moving to for a better recording place]

Interviewer: Perhaps you can summarise what you were saying about how you use diagrams in communicating your research?

Participant: Ok so in summary, I use a diagram when I want to summarise or represent a higher level view of some process, or the building blocks of a method that I’m trying to convey. In the case of neural networks, usually when I want to show the building blocks of the network, inputs, outputs, I draw different layers and how the connect and so on. When it comes to networks, we need that in order to simplify the representation of the network, for the person reading to get an idea of the architecture or the general [unintelligible].

Interviewer: Why do you use a diagram to do this rather than text?

Participant: So [Pause]. Point 1 is you sometimes need a lengthy description, and it is better summarising this in a picture. The second one is sometimes the text is ambiguous, in the sense that, especially in scientific writing, and particularly in CS, there are terms that are used in different ways across the field and related fields like statistics and mathematics. Sometimes the picture disambiguates from these terms. I don’t think we need to have some standardisation regarding diagrams, but [unintelligible]. When I look into a paper about a certain subject [unintelligible]. It is easier on the eye. Reading a long paper of text, sometimes it is tiring, and you want a picture or graph. I guess that personally, when I read a paper, I want a nice visual representation. I can better memorise it, and other people are doing the same thing I guess. That’s about it, for me.

Interviewer: How do you decide what to include in your diagrams?

Participant: It depends a lot on what kind of idea I am incorporating in the text. In a general way, I would. Can you repeat the question?

Interviewer: How do you decide what to include in your diagrams?

Participant: In a general way, I have something which is novel, and want to express the main concepts which differentiate it from the prior art, from the literature. I want to help others to identify the prior art. So there are those that are expected. I use the diagram to highlight the points of the prior art, from the literature. The rest is connecting those main ideas that I’ve expressed to the general principle, and things discussed in this paper. This is the general point of identifying what we need to do. And the rest is connecting.

Interviewer: So how do you decide how to represent these things?

Participant: That’s a good question. If the thing in question is going to be connected to other things or has some kind of graph representation, it will be squares or circles or some kind of simplistic object with words written on them. Where there are standard representations for them I would use those, for example like a UML diagram. I would do something like that for an ER diagram, because of the standard elements of diagrams. Apart from that, basic squares, circles and lines, with I’d say minimal colour is great. [unintelligible] Was there an interruption?

Interviewer: Yes temporarily

Participant: Ok so there are these basic elements, and I’d use standard elements if I can. If I need to represent some kind of operation or something like that I might find a more creative way of drawing [emphasised] that, in a block or something like that. But yes, usually this is the way I build these representations.

Interviewer: So the ones which have become conventional, that’s fine, do you copy from people that are doing novel things as well, or is it only if it is conventional?

Participant: Yes, oh yes. It depends on what kind of novelty, but usually when there is some kind of diagram or specific way of drawing that, it can provide some sort of inspiration. If there is some type of block, on a certain subject, I would certainly use it, it helps to comprehend it if you see what you are expecting to see. So after seeing one or two times a certain type of visual representation, I start to expect to see that, and it basically becomes a kind of a standard. So yes, I certainly I do this. I choose things that have been useful to me either the idea that is being expressed in the paper or even the results, if there is a nice way to plot or visually represent the results of the experiments. If that is shown to be useful, if it helps me I would certainly use it in my own work.

Interviewer: So how do you decide what level of granularity to go to in a system architecture diagram?

Participant: Ok so it depends on the length of the paper. If it is a long paper, with a lot of space to discuss, so I can get into details, I will use a detailed diagram, I will bring a representation over several figures in the paper, showing in detail the thing I am trying to express. In a short paper, I will not have this much space to go into detail, and I will try to condense this representation to summarise the general idea of the process, of the method. Or if this is my method processing pipeline, I do this this and that, I’ll summarise it. Or in the case of NN I will show the modeling, it depends on the space available if I can go into detail or not. This is the basic step to decide the level of granularity. But apart from that, there are some papers where to understand what exactly is happening in for example a certain NN architecture, a certain technique that is being used, even if it is a short paper you need to go into some level of detail. In those cases usually I am going into detail on the thing I want to highlight, and complement it with explaining for each part of the picture, what is happening there, and so on. This balance between text and the figure, the detail of the figure, depends a lot on the need to show certain things and of course on the space available. But this is secondary. First the space available, and if it needs a detailed discussion or not. And second how do I need to highlight this particular part, and how to balance with the text.

Interviewer: What software or tools do you use, and why?

Participant: Usually, depending on the type of diagram. For basic stuff I use these days openoffice draw, because it is very flexible in terms of formats, and it is easy to draw basic shapes and connect them. I don’t have to worry about colouring scheme, and it is pretty good in this aspect. And it’s free, I don’t have to pay anything for it. I have it installed on all my machines. It also allows me to export, it is flexible for exporting vector formats, pdfs for posters, or svg to put on the web, or whatever template the paper is working with, LaTeX and so on. So yeah, for basic stuff openoffice draw. In some cases I want to do something more sophisticated, and I use inkscape. Also a free software, it is a bit more restrictive regarding format. It accepts them, and allows me to do some operations that I would not be able to do in Draw. And in some cases if the pictures are very simple to edit I would use picture bitmap. When I want to draw the diagram, I want to lay out in a very specific way, have the colours and so on, something very manual. But if the diagram I want to draw is something that is derived from a result or something I want to automate, and can be expressed as a graph, I use the graphvis system, visually in the form of a graph dot file, which can be processed by the dot program to generate a picture graph diagram. Then this can be used for a presentation or anything. Also this utility allows me to generate both raster and vector graphics. For automated diagram generation that is very good. In exceptional cases, if I’m working with UML diagrams that I want to share with other people or somehow collaborate with others, I would use the astah software. That is specifically for working with UML diagrams and the like.

Interviewer: Thanks. So moving on to how you use diagrams when you read a paper?

Participant: Basically in the same way as I write it, but in the reverse. I am looking for those expected elements, if I’m looking for a paper that explains say NLP system pipeline, I would expect to see some building blocks of that pipeline, for example this part does this processing which effects this using this kind of technique, and so on and so on. This is the kind of thing I would expect from the diagram. In the case of the NN I would expect the buildings blocks, the functions usually involved, and some external elements like this subsystem generates the inputs, this formulae does this and that. This is the first thing I am looking for when I’m looking at a diagram.

Interviewer: Why do you want that? Is it the same as when you are writing it?

Participant: [Misunderstood or misheard the question] It helps me comprehend what the author has tried to do, instead of the text. Sometimes reading the text, they have a textual explanation, and they point to the diagram. And then I look at the diagram to see if that is actually what they’re saying, or if it is a different idea, especially for example for system architecture diagrams, sometimes the basic building blocks are what you expect they are, but sometimes the connections are not. Sometimes there are differences that you can get insight from that that you didn’t get from the text. Sometimes it is not as easy, or it is ambiguous, in some kind of explanation. Probably also with some terms that are ambiguous among the different fields of computer science, the diagrams are a good place to understand that. This layout here ah this architecture here is laid out like that so this is what it was trying to do. Or sometimes you think you have an understanding, and when you read the diagram it is not particularly correct, and you think “oh, what is it doing here”, and it is doing something else that is not quite matching what I expected given this textual explanation or so on. But it is the place we look for, or at least I look for, a better understanding of the explanation. For memorising, because it is a picture of it, so I like that, as a way of removing doubt about the system, or having some kind of confirmation.

Interviewer: I’m aware of time, shall we move to the diagram examples? A

Participant: So this gives me the basic blocks of the network, and that’s something I would expect to see, like the input nodes. The layers here and probably some idea that is expressed in the textual part, regarding this block. There are some blocks where some kind of score is calculated, and it goes to softmax and so on. So what I like about this is that it is following a certain type of a standard of neural networks, this which is in the input blocks, and any kind of input data, the embedding layer and so on, in the bottom, and the network goes upward where we see the things happening. So this facilitates me understanding the network is doing this and that. What I didn’t like is that there are some things that are not very well explained here, for example this merge. That has a black and white or yellow circles on it. This is the kind of thing that if I looked at the textual explanation would be better, but this is the kind of thing that unexplained doesn’t help. It would be better explained in the caption or some kind of legend. For example, when you use this kind of colouring scheme, it is nice to have some kind of legend to help people’s understanding of what is the data that is flowing through the network. We’re missing a legend from this particular diagram, the rest is fine.

Interviewer: So overall do you like it or not like it?

Participant: I do like it.

Participant: Diagram B is quite high level. I see the building blocks of a CNN, apparently. You have convolution here, and a siamese network. Ok so this one is very high level, it basically shows us that the general idea here is about sharing these layers from this FCN after a certain point, and using that embedding space. What I like about this is it is very simple, and very high level. It is easy to get a general idea of what is going on here. Also, the caption explains this general idea and gives the meaning. It is quite good for the purposes of conveying the general idea of what the model is doing, but not for explaining the NN architecture. So what I like is the general idea, but I would not use it for explaining the NN architecture.

Participant: So for diagram C. [Pause] This is filtering, ah ok. This is a system. Ok. The method. System architecture. They took images from here. Filtering. Ok. I don’t really like this figure because if I am not very aware of what is going on here I would not understand the meaning, especially of the images in the bottom left, they are very specific to the method here. I think this does not do well as an architecture diagram. The elements are here, the inputs and there are actually no outputs here, just intermediate processing. I don’t like this, I don’t think this gives a better idea of what is going on here unless I am really aware of this method.

Interviewer: Just on this one, do you have thoughts on the use of the example throughout the diagram?

Participant: You mean the picture. Yes, I think it would be better broken down into smaller, smaller in the sense of detail, figures where each one is a step of the process. It is condensed so I cannot see the detail, but the idea of a step by step chain is nice.

Participant: For D. [Pause] This is an encoder, ok. [Pause] Ok so this one. It’s a step by step encoder with a structure that are used to proceed with the encoding method. I kind of like it, it gives an idea of what it is and what is happening at each step of the encoder, how the memory is referred to. It would be nice to have a clear output here, but I guess you get an idea of what is happening here, even without knowing the contents of this work. Let me check the caption [Pause]. Yeah the caption is good, I definitely want more detail. Yeah this D is nice. I like all the parts of it. I would put something else here for the output, but overall it is quite good.

Interviewer: So for letter E?

Participant: Ok. This one is a general architecture. I think it could be very well for a poster. I communicates the idea very well, identifying all the building elements that I would expect in this kind of system. There are some things I am not so familiar with, in the NLP area. This is nice yeah. I would prefer a better caption on it, but yeah, this one is good.

Interviewer: On this one, things like the Context Aware Fusion module, we have 10 rows, do you think this is significant? And similarly for the number of columns in the feature matrix?

Participant: No. I don’t think they are significant. I would not look for significance on that because the pictures, especially because in this case with the general architecture, are not the place to get this kind of detail in terms of a specific amount of nodes and so on. This is general figure for architecture, so no I don’t see or I would not look for significance in the number of rows.

Interviewer: Things like hyperparameters, you don’t want those in diagrams?

Participant: Usually no, because they are usually very specific to experimental setup. Usually they are adjusted. If you want to see that in an experimental section, they are not fixed. I would not expect to see that in an architecture diagram, certainly not.

Interviewer: In transcript embedding, there are self attention 1 and n, and they’ve omitted the ones between. Does that bother you or is it fine?

Participant: That’s fine. Sometimes I would see three dots or some kind of notation of this kind, but if I can clearly see there is some kind of structure in between and there is clear indication 1 and n that there is something in between then that is fine. It is very understandable.

Interviewer: A lot of this iconography, you could use half the space by removing some of the rectangles and circles, and lines and arrows. Are you comfortable with the amount of stuff there is in this picture?

Participant: Yeah, initially I said this would go well in a poster because I think it would look nice in a poster. It is fine to convey this idea into a paper. I would break this diagram into smaller pieces, and in those smaller pieces I would use much less geometry. In this combined diagram it would use less space. I understand in this particular diagram this was done not for any specific technical reason, but because it makes it into these nice rectangular shapes. It is easy on the eye. That’s why I say it would be good for a poster, this being a big picture makes it easy to grasp and handle. You get the idea of each point. If I wanted to take smaller space, I would break this up. For example Context Aware Fusion module, I would have at most 3 or 4 rows with three dot notation, same for procedure extraction module. This transcription embedding part here is the only one that I think I would draw the same way. The rest I would certainly change, a bit more for space reasons. But I understand why it was drawn like this. When I do posters, I tend to do the same, and draw it a bit into a specific shape for the thing I want to show. When looking at a poster, you want people to look at the poster in a certain way so the easier it is for the eyes to follow a certain path on the poster, the easier it is for them to get the information. You can have a bit more and add lines to do this. For example in procedure captioning module, I’ve done a very similar in the past for a poster, because it is the kind of thing that I have this space and I want to point at each part, and I have a lot of space to draw lines. When looking at a certain distance they can get the general idea, and when they get closer to the poster they can get into the details. I think it is a poster thing, I would certainly not do that for a paper, because you’d almost certainly run out of paper.

Interviewer: F?

Participant: Summary of. This is pretty much the usual kind of thing that I do when I want to explain some sort of pipeline. It is at a very high level, and usually you want to highlight the basic structure of your method or technique. Usually I would then do some colour on the boxes to highlight, and the connections and so on. Usually this diagrammatic thing that I’m trying to do is to express exactly the same thing as is expressed here. I am suspected [sic] to say I think it is good, and I would certainly do the same kind of drawing when representing this sort of thing. Unless the discussion asks for some more detail. Then, usually I would draw one of those, and then proceed to break it down into what this block is doing, and then make another diagram for that specific block, going into detail.

Interviewer: So this contains enough information to be useful to you?

Participant: In this specific instance yes. What I’m looking for is the main building blocks of whatever the technique is, so of course I would in most cases expect more detail. For example in a 10 page paper that is discussing a single method and I only find this one, I would be looking for more information because this would not be enough. For a short paper, this is ok. This gives a very high level, a kind of overview of the structure, but would certainly be better complemented by detailing some of those blocks. Or in this case here, the projection layer and bilstm are very standard, at least. I know what to expect from them. The mean over time needs a bit more detail, and the FC layer is fairly standard, everyone is using this kind of network. I would be looking for more detail, depending on the size of the paper, on the projection layer and the mean over time, and maybe about the score, unless this was something that was already explained in the text. It probably was, just looking at this instance here. In terms of diagram, I would only be looking at more detail for the projection layer and the mean over time. The rest is those kind of blocks that I would be expecting to see in this kind of study.

Interviewer: We’ll go as fast as we can, moving to the task list. If you could tell me the 2 or 3 that you think are most important in your use of diagrams, and any you definitely don’t do.

Participant: The first, certainly. Identify novelty or contribution, because basica building blocks I want to find. For example the last diagram, this is pretty standard, but what is different, what are the highlights here. This is what I am trying to get from the diagram. So the first one, certainly I would rank it very high. Identify layers, yes. This is also a basic thing I would look for when looking at a diagram. External dependencies, sometimes. They are not essential to me, they only make sense if you have a system architecture kind of diagram, I would not be looking at them in general. So not too high. Corpora and data types. Data types sometimes, corpora never. Because corpora it should usually be explained in the text. It does not go very well in the diagram, and most people put a little box saying this is the data and this is the corpora used, I don’t look at the diagram for this. Data types they help you understand what is going on. This would rank very low and relatively high. Identifying representational choices, this is part of it. This is a part of the diagram, this is a UML diagram, and so on. So yes, the choice of representation is certainly very important. The purpose of the system, no. For this I would be looking in the abstract. From the diagram I want a basic description of the system, that is correct for representing your idea in terms of how significant and how it connects, not about the purpose. Arch features, very high, second place after. First two are 1 and 2, features are 3. Opportunities to alter, certainly yes. If working in the same field you are looking to compare to change something. I would put that 4th or 5th. Identifying important, yes this kind of relates to arch features, because usually they try to highlight something. Comparing to other systems, certainly, related to altering. Initial check, yes, when I’m looking for specific elements, if one of those elements is highlighted by the author that is certainly. 4th or 5th place. Index to navigate, certainly not. Memory aid, definitely. Summary aid, no. Understanding how, this relates to almost all the ones that I have ranked here. 1 novelty, 2 2nd, 3rd arch, 4th check, 5th alter, 6th author, 7th memory. The rest are ranked very low

Interviewer: I might have missed it in the very beginning of our talk, but when you are reading a paper, how do you use the diagram? Do you turn to it first?

Participant: Using this list here, I can say I would first identify the layers and then

Interviewer: No I meant in the context of a paper. Could you describe how you read a paper, what order you do it in?

Participant: Ah, in the paper. I would read the paper in the usual order, starting from the abstract, to the introduction, and so on, in the textual order. I use the diagram to support the method understanding. As I said before, I usually look at the textual explanation and then look at the diagram to support what I am understanding from reading. When I look at the diagram, I look for the things that I mentioned now, in that list.

Interviewer: Super. The last part is a little set. Do you use diagrams as part of your creative process when doing research?

Participant: It depends on the subject. When I’m working with a systems architecture, yes I do. When I’m working on say an optimisation, sometimes you need to, especially if you are working in a group, you need to share your ideas in a group. We’ve trying to do this and that, and diagrams are a good choice to do that and then discuss some sort of idea referring to it. Before and during building something, it is good to have the aid of diagrams to get people to discuss. In some lines of work, like when I am trying to optimise or introduce a new process, it may not be aided by diagrams. It depends on what I’m doing.

Interviewer: Are there any tools that would help you in making or consuming diagrams?

Participant: For making, there are the tools I mentioned, which I use. These days I am looking more into dynamic diagrams for things like decision trees or even in NNs sometimes there are ways we can better represent them in a dynamic way, like how the network operates over time. I am actually, these days I am looking into that, as I am still not using such tools. I know they exist but I do not have the knowledge to take advantage of them, at the moment. So that is one thing.

Interviewer: What about, have you used tensorboard?

Participant: My experience with tensorboard is not much, but it has helped check progression, how the network is converging, something like that. But in terms of thinking more about the results, this is the current setup, this is how it is all organised, I think this is unclear. Maybe this is a lack of my knowledge. I think it is useful in a situation where you want to talk about convergence graphs or something. I don’t see me using this most of the time. It may be a lack of knowledge.

Interviewer: I was just interested in what your experience was. You seem to quite like the block diagrams, the type you would make and the last example, do you think this very schematic way of drawing, maybe with some colour, would you hope that everyone would use this way of communicating, diagrammatically?

Participant: For that kind of subject, yes, because it is simple. I can look at it very quickly. The colour aspect is not an important thing for me, but sometimes it helps. I like to use colour and in some diagrams it helps to express an idea or to highlight something, usually with a legend, which was one problem we saw. I’ve been seeing these kind of diagrams for a while now and they help to summarise the idea of the architecture, give you an idea of what to look for, and it is the simplicity that helps to check this explanation in the text. “Yeah ok this is what they are doing”. So the more complex the diagram is, the more time you need to spend to understand, especially if it is drawn in a non-conventional way, because then you will spend some time to check the elements that you expect to find there, and try to interpret the visual representation in the icons and so on. So yes, the basic building blocks and squares are usually enough to pass the idea.

Interviewer: Anything else you can think of that might be interesting to talk about on this?

Participant: Yes. I just want to complement some things about what I am expecting to see in terms of things that I would expect to help. We are still very used to seeing static figures and so on, but many of the works these days deal with dynamic systems, and deal with things in a different time in a different way. So I would like to see some sort of animation, so those pictures I’ve been using recently with visual representations that try to show how a NN does it process, how the nodes are changing over time in a visual way. If this can be done with simplicity, that would certainly help with comprehension and developing the subject in question. I think I would like to see more, especially now with many papers published on the web, I think we can take advantage of this resource.

# Participant 9 (P9)

Interviewer: How would you describe your experience level with CV and NLP?

Participant: I’d say certainly not beginner and not detailed expert. I’d put myself somewhere in the middle, on the practitioner side of things. Not building any networks or anything from scratch, but getting the existing networks to actually do something, and interpreting the output.

Interviewer: You’re in industry would you say?

Participant: I don’t think I can say that, because in no way is there profit. I’m employed by the university, it’s not for profit. It is to underpin research, so I would say not industry. We have this sector called “academic related”, which is not primary research, but the supporter of a primary researcher. Primary researchers pose the research question, and I make the technical stuff happen which will allow them to address the research question.

Interviewer: There are a couple of questions, and a couple of activities. The first question is can you describe how you use diagrams when communicating your research? I appreciate that you may not have research to communicate, but in general, how do you use diagrams to communicate. Diagrams in general, but with a focus on NNs and system diagrams.

Participant: That’s hard to answer. I suppose I use them quite a lot in the other branch of my work, which is about distributed analysis. It helps to have a picture of where different bits of processing are happening, and how data stays put and analysis commands move around and results come back. You can say what is actually happening, but it helps if you draw it out and it makes more sense. That would be an example in the general sense. For the neural networks, I think it is rare that I would need to explain to anyone, in any sense, the detail of what is going on. I might throw up a generic picture of a CNN, but not really explain how it works or what it is. Almost as a background, if that makes sense. But actually I did do a workshop the other work, and did explain at a high level how a NN might process an image, so I put a diagram off a website with 4 layers with the classic circles representing activation. With input layer, lines for pixels in the image going to the input layer. Then two hidden layers, with various connections to the input layer, and an output layer. And it was doing the NMIST digits with 10 outputs. I explained it first of all by a black box with just layers and connections, and saying basically you have have some complex input you want to make a prediction on. In this case, you have an image, which is a complex input, and you have something you want to predict, the number 1 to 10. Then I’d explain the process, the black box having a complex set of functions in there, with loads of parameters. You can imagine that by setting the parameters you can get something you know is correct. How you set those parameters is by training, so you feed it an image that we know the answer to and we use that to set the parameters. Then I took away the black box and explained how a node is activated, and the inputs and threshold cascading through. I used a sort of detailed but simplified diagram to explain to people what these things are. And that is quite indicative of how I use these! [Laugh]

Interviewer: So you don’t use them when you’re trying to make something? You don’t sketch?

Participant: No, because at the level of practitioner, by using the tools that I do, a lot of that is done for you. I literally go “this is my input data, it’s a 256 by 256 image, I want resnet-50 and I have 10 output classes” and it does everything for you. It’s almost taken the sketching out of it.

Interviewer: So is it fair to say you are focused on inputs and outputs, and the stuff in the middle is a lot less important?

Participant: For the purposes of doing work, yes. We’re not in the business, we’re not in the research of designing new architectures. We just want to take the best architectures out there, and get it to work for our problem space. So I don’t ever, or maybe I have occasionally done sketches, but only for my own benefit to understand what is going on. And quite often if I want to know, I will just probe what a particular set of activations might look like, and I’d do that by looking at what dimensions the tensor has, rather than drawing the thing end to end. So I’m afraid I don’t draw much, maybe I should do more.

Interviewer: The kinds of things you do involves taking existing state of the art, selecting something, and applying it, is that right?

Participant: Yes, or dare I say it, borrowing someone else’s code, where they’ve already done the difficult stuff, just because of the time constraints. Why would I write something from scratch when I can use existing stuff and get it to work.

Interviewer: So not really building new pipelines, as it were?

Participant: From a 256 by 256, doing 10 convolutions on it, 128 by 128. No, I am just taking whatever has been found to be good and reusing that, I’m afraid.

Interviewer: So, what tools do you use to do this?

Participant: "fast.ai" is the framework, built on pytorch. Typically you are working on a jupyter notebook, running it on the HPC. Interactive or batch processing.

Interviewer: What about when consuming research? How do you use diagrams?

Participant: I guess just in order to get a broad understanding of what is going on, rather than understanding in detail exactly what is happening at different points in the pipeline. Again, as we’ve touched on, it isn’t that I don’t want to understand, but it doesn’t necessarily help you getting the thing up and running. I’d say one of the vision tasks, I was trying to use retina net, which is like a U-shaped thing, network, pipeline, whatever you want to call it. On one side you have your resnet backbone, so you go down the whole image down to something with feature maps at each level and at the bottom it is maybe 16 by 16 pixels, and you reconstruct up the other side, with connections across where you build it back up. I struggled to understand how that is actually working. But by looking at the diagram, it made me realise why it was not working in the area I was applying it, because it has some very small objects. The actual architecture constrained what size you could have. So yeah, those times are rare. It is quite bad, but I don’t read that many papers. I don’t need to, because someone else has already done that and put it into the framework.

Interviewer: So it is directly from the framework that you learn about techniques, or is it from blog posts or something else?

Participant: Normally from this "fast.ai" course, and the related posts in the forum. In the course they will do a certain amount of stuff, and on the forum people say “I tried to do this, here is my notebook” so I might use that.

Interviewer: Ok, yep. Let’s move on to the examples. If you just tell me what you like and don’t like in these diagrams. [Time pressure due to people outside meeting room]

Participant: Ok so A, I guess I quite like that one because it has skipped out the details and the precise dimensions of what’s going on here, and you can pick out quite nicely what is going on. But it is hard because you don’t know what x is, is that a word, or the whole text, or what. It seems quite high level, but hard to know exactly what is going on.

Interviewer: In situations where you’re reading a paper, x is probably in the text somewhere, would you prefer it to be in the diagram so you don’t have to look it up?

Participant: Yes, or at least in the caption, or some sort of key, something like that. I think it is hard to know, I don’t know what score is. The other things I know, but those need clarification.

Interviewer: Next one? B.

Participant: [B Pause] I guess this one feels relatively familiar, as it looks like two convolutional networks, but I don’t know what FCN is, so that might need clarification. Running in parallel. Then I don’t really know what’s happening in the end bit. The embedding space, and what contrastive loss is. I guess it is hard to disambiguate between what I don’t understand because I don’t know, and what is good and bad about the diagram. Sorry about that. Also I am not clear, in the two streams, there is an image, and a silhuette, I’m not sure what that is. There is a bit that feeds into the green and a bit that feeds into the yellow, I don’t get what the one underneath is. But I never looked at a network like this.

Interviewer: Overall, do you like or not like?

Participant: I’d say not like, but it is hard to say. I guess it doesn’t make it clear to me what is happening, but that is coming from an uneducated point of view, I think. That’s ok.

Interviewer: Next one? C

Participant: Ok to me this looks like nothing I have ever seen before. I don’t like it. [Laughs]. I have no idea what it is doing. That’s my answer to that one. I’d need to go and research a load of stuff before I had any clue on this one. D, this one also looks like nothing I am familiar with. It feels a bit RNN-y, a bit like a recurrent network of some kind. At least the concept of encoding is familiar. I don’t really like it because I don’t really get what’s going on. It’s hard, maybe if I had looked at these in advance, but I would probably have to spend a couple of hours working out what these are. Sorry.

Interviewer: Do you think you would be able to do it with just the diagram, or would you want the paper at the same time?

Participant: I think I would want to read the paper at the same time, and then probably look up a load of other stuff, because the paper would probably assume more knowledge than I have in this domain. I promise I have read an ACL paper [Laughs] but it was one mentioned in the course, so it felt quite cosy and familiar.

Interviewer: I’m pretty sure you are part of a large group of people like you, if that makes sense. E?

Participant: Oh goodness. At least it is obvious what’s going on here. [Pause]I’d like to say it is somehow taking both a video and a text and using that to tag each frame in the video, or at least each segment. As to what’s going on in the detail, I’m a bit lost. Text and image and bringing them together. It’s a bit beyond me. I’m sorry.

Interviewer: Don’t apologise! Are you using the examples in these to try to understand the purpose and how it is working, if you see the distinction?

Participant: I think seeing what the actual input is, is the most easily digestible bit. You are immediately drawn to the inputs to guess what is happening in the rest of it. And then looking at the output helps to guess it is summarising the transcript. But then the actual bits of the network seem quite high level, so it’s a bit odd, if that makes sense.

Interviewer: Is familiarity something you’re looking for in terms of both the content and the visual encoding of that?

Participant: Well, it has to be otherwise it is just a bit alien. It is so complicated, or feels complicated, unless it is familiar I am at a bit of a loss as to what is going on. I can see there is some sort of resnet-34 going on there, and some sort of concatenation going on there between the visual bit and the text bit.

Interviewer: Overall do you like or not like this one?

Participant: It’s fine but it’s a don’t like.

Interviewer: And the last one.

Participant: It’s quite nice and simple, with each block. It hides lots of detail. I think that once again, with the inputs and outputs, it looks like somehow you’re putting in a summary and a full text, and generating the score, but I don’t know the score of what. I don’t know. Maybe they have a lot of examples of good summaries and bad summaries and they know what a good summary looks like. They have BiLSTM, and merge. The two branches, I’m not that familiar with, because the tools that I use don’t tend to take that approach. I like it overall.

Interviewer: I’m aware of time, so the last part is a task list. What’s helpful is if you can pick out the top 2 or 3 most important in your use of diagrams, and any you don’t do.

Participant: Purpose. Types of data and corpora. Not altering the architecture, because it is not what I do, that’s too difficult. [Pause] I guess the author part, about why they think it is special because that will help me identify if it is best practice or not. I think that’s all of them.

Interviewer: Is there anything that would help you in consuming diagrams, any tools?

Participant: I think it would help to have a tool to help understand diagrams, or a tutorial on reading types of diagrams that are out there. It was introduced in the course but there are slight variations, and there are no standards. It would be nice to have a tutorial on that, or a guide.

Interviewer: [Time finished] Anything else about diagrams?

Participant: I wish I did understand them better, because I think that would really help. If I had time to sit down and really work through them. I can get away without doing this, so I don’t.

# Participant 10 (P10)

[This participant often spoke in rhetorical questions, this is not a transcription error.]

Interviewer: How would you describe your experience level with NLP and CV?

Participant: Do you have a scale for that, or should I answer freeform?

Interviewer: Freeform.

Participant: Natural language processing, surface level, attended a few talks, read a few papers. I’m very good friends with one of the I think leading researchers in the area, a guy called [Redacted for anonymity]. We talk quite a lot about his research. Computer vision, I am quite familiar with traditional CV techniques used for the last 10 or 20 years, all the classical techniques. Fairly familiar with more modern techniques, implemented a few toy projects or proof of concepts. I’ve never deployed either area in a production setting. Does that cover?

Interviewer: Yes, so which is your stronger suit?

Participant: Probably CV, because I’ve done more with it. With NLP, I’ve just read about it, and thought about it.

Interviewer: So have you coded stuff up?

Participant: With CV, yes. Like I say, all the way from classical stuff back in university ages ago, up to more modern NN style techniques, and stuff in between.

Interviewer: Is this part of your job?

Participant: Yes, it is a little bit of my job, the proof of concept side of things. It’s part of it. I spend my time doing several different things at the moment. Part of that is implementing proof of concepts, or investigating whether it is possible to do something which would help. This isn’t necessarily CV, it is anything data related. A bit part of that is of course visual, and a big part of what we do is deal with video, so understanding what we can do with that is quite interesting.

Interviewer: Thanks. Some of these questions are more geared towards academics doing research, and contributing in that way, so if they don’t make sense we can skip over. The first question is about how you use diagrams when communicating your research, and I suppose that would also include industrial output in some sense.

Participant: Sure. [Pause]. Are you asking about system diagrams?

Interviewer: Diagrams in general, diagrams of NN systems and their systems architecture are particularly interesting, but any time that you are using diagrams is interesting.

Participant: Ok. So as you know diagrams are extremely important for conveying information, blah blah blah. What tends to happen in my job is we tend to develop a proof of concept and show the outputs of that. If we’re doing image processing stuff, it will usually be a couple of things, here are some of the example results. If we’re using image processing to overlay something we’ll show the results of the overlay, so bounded box, segmentation, that sort of thing. Then I’ll also use diagrams for displaying confidence levels or effectiveness at distinguishing features, classifications, and those tend to be fairly straightforward charts, graphs and probability measures. Fairly straightforward stuff. System level things tend to be more abstract. I’ve never had to talk about detailed novel NN implementations, so it has all been fairly high level block diagrams.

Interviewer: So when you’re communicating the systems as block diagrams, why do you use a diagram to do that, rather than text?

Participant: Umm. I’ll blurt out a few things and we can refine. It’s a well understood medium. Computer scientists get it. Directed, not necessarily directed, but graphs in general are understood by computer scientists. It is a commonly used format. It encodes a couple of things quite well, it encodes data flow and also computation steps, so it is a nice way of doing both of those two things at the same time. [Pause] It is generally easier to walk through people, so if you are presenting some design it tends to be easier to walk people through what is going on by pointing at blocks and describing that particular block in the overall picture.

Interviewer: Sounds good. So what sorts of problems do you encounter when you are creating these systems diagrams?

Participant: [Pause] It tends to be a tradeoff of complexity vs information transfer, I suppose. I tend to, for example I may have two block diagrams, one with more detailed information than the other. So I’ll start off with the more basic, high level diagram, and as the explanation carries on, as we go into more deep areas, I’ll bring out another block diagram with extra detailed filled in. I suppose that is the most obvious first one. It is not always clear where that line is drawn, how to choose the level of complexity. [Pause] I suppose because they are so general, you always have to do some explanation of what is going on. I suppose for example a state machine is a very common thing, so people usually get it pretty quickly. For less common, for anything else you have to explain what the blocks mean and what the edges represent etc. What other problems? I suppose drawing them can be a problem. If I can, I’ll generate them automatically from whatever I’m doing, and use something like graphviz dot dot format to describe the graph and then generate the png. It doesn’t always work. What other problems? [Pause] I think those are the main ones.

Interviewer: The complexity to information transfer is super interesting, and this idea of having two diagrams to communicate at different levels of granularity. So when you’re doing that initial drawing, is that covering the entire system, the entire dataflow, inputs, outputs and some parts between?

Participant: Yes, I would say so. I would start with the high level diagram that shows all of the functionally interesting pieces. The pieces you want to talk about because they are something essential, or they contribute something essential to it to light, and I would leave out the implementation details, if you like. The intermediate steps that are not particularly important.

Interviewer: Is this something different to have input, preprocessing, make some model, do some loss, get an output? Is it more granular or specific than that?

Participant: I think I would probably make it slightly more specific, but not necessarily so. [Pause] What’s a good way of answering this? [Pause] I mean people get the general flow that is going to be there in any sort of data science project or data processing project, that they expect there to be get the data, preprocess, clean, train, whatever, so I guess what is more interesting is what are the essential parts of that pipeline for the specific project I am working on. So I’ll probably dive straight into what does the train block look like for this dataset, or for this model. Or for the model I would dive in straight away and say what does the model look like, what is the processing, what are the key elements of the model. And similarly for validation, what are the key points in the validation. I suppose if I was presenting to someone new to this, and I haven’t really had to do this, if I was presenting to someone new to the field I would show them the top level game plan of data in, results out, with papers in between.

Interviewer: So your use case is you tend to be explaining to someone who already knows a bunch about it?

Participant: Yeah.

Interviewer: And you include training and evaluation as part of this systems diagram?

Participant: I would actually do those separately, because like I say, there is implied there is this overall system.

Interviewer: You spoke about autogenerating from code when you can, are there any other tools that you use to make diagrams?

Participant: [Laugh] A little bit of Visio, for my sins. [Pause] And that’s it.

Interviewer: Do you use pen and paper and whiteboard, just out of interest?

Participant: Yes, sure, absolutely, that didn’t even pop into my head. Yes both of those absolutely. I was thinking computer tools.

Interviewer: So was I, but I was just interested, because one shouldn’t make assumptions.

Participant: Yes, absolutely, pen and paper is the first place to start. And keep returning to.

Interviewer: Sure. So we’ll come to this a bit later, some of this. How do you decide how to represent the thing? Are you dealing solely in block diagrams, or do you do an iconographic things?

Participant: [Pause]. Yes. [Pause] Just thinking about how to answer this. Definitely, possibly a slightly different area where we use a lot more icons is web-based systems. When we’re talking about deploying or figuring out how to deploy a model into a production environment then we’d be describing cloud systems with lots of icons and block diagrams, but mainly icons. For the ai and machine learning stuff, I suppose the generic icons that you get that you see all over the place, but nothing particularly special there.

Interviewer: When you say the generic icons you see all over the place, what kinds of things?

Participant: Like if I wanted to represent a sigmoid, I’d have a little icon with a sigmoid. Or like convolution stacks.

Interviewer: So how do you use diagrams when consuming research?

Participant: They usually are the first thing. When I read a paper, I’ll usually read the abstract, scroll through quickly, sometimes jump to the conclusion, and then look at the pretty pictures. Then what do I do. If I can’t immediately understand what a particular diagram is doing, I’ll try to read some of the text around it. But first thing is can I get enough understanding of the whole paper by looking at this diagram. That’s the biggest use case actually.

Interviewer: When you say reading the text around it, do you mean that literally?

Participant: So yeah, if I look at a diagram and I really can’t figure out what this is doing, I’ll usually read the paragraph before, or I’ll scan the text around it to see where the figure is referenced and read that bit. I don’t have enough time to read whole papers.

Interviewer: Do you have any problems with consuming research more broadly?

Participant: Ooph. Time. There is just so much stuff going on in the field that it is totally impossible to keep up with everything. I suppose that is a big problem. Figuring out which papers to read, what to look into. That’s a broad answer.

Interviewer: How do you find what papers to read in the first instance?

Participant: Paperswithcode. I really enjoy paperswithcode. A lot of paperswithcode, and then I browse arXiv basically. I don’t spend much time in journals, mainly because I don’t have subscriptions, and I find that arXiv has more than enough interesting stuff. The batch newsletter as well. Occasionally stuff on nvidia's website is interesting. What else? Hacker news, although that tends to not have the latest interesting stuff. [Pause]

Interviewer: You wouldn’t look at conference proceedings?

Participant: I was literally about to say conferences. Conferences. The obvious one you don’t think of first.

Interviewer: Quite a lot of what you’re describing is not being curated by ACL or something.

Participant: Yes, for some reason I don’t like curated stuff, I prefer uncurated stuff because it tends to be a bit more raw.

Interviewer: And therefore?

Participant: Less biased. That might be a bad thing from myself, but I prefer it.

Interviewer: Cool, shall we move on to some examples. This is a series of 6 diagrams, and if you could tell me what you like and don’t like about it.

Participant: Ok. Diagram A. [A Pause] I don’t particularly like having to turn my head to read the labels. How nit-picky do you want?

Interviewer: That is great, that feedback is fair and good.

Participant: Yeah, I find that really annoying actually. Colours, good. [Pause]. I’m assuming that is just doing concatenation at the top for the sigmoid. [Pause] Yeah, ok, I suppose this always makes more sense if you have an equation to match the diagram to. I can kind of guess at what’s from the embedding layer, I can guess what G is doing, but it is not entirely obvious from the diagram why. But that’s fine. Dotted line, ok. Yeah so I suppose broadly it makes sense, I can see what’s going on. We have a bunch of vectors coming in and they are doing through a scoring function, being combined with this embedding vector G first, getting softmax on all of those. Yeah. So presumably that’s just a normalisation, but it is not entirely clear. Yeah I don’t know what’s going on with the three black, the input to the sigmoid. It’s not entirely obvious to me what is going on there. Merge classification. Sorry, merge then classification. [Pause] Meh.

Interviewer: Overall do you like it or not like it?

Participant: Hard to say without a scale.

Interviewer: We can come back and do at the end?

Participant: I guess I like how signals are being labelled, it’s using conventional dots, I get what the dots mean. Typography, meh, I’m not a massive fan. Don’t necessarily know what the different colours mean above the dense layers, but maybe that is just to make it look pretty. That’s cool, colours are good. Do I like it? The issue with it is probably the side labels.

Interviewer: I’d like to follow up with a couple of specific questions. The four circles and the five circles, do you think that represents anything?

Participant: I hadn’t even spotted that. So we’ve got the embedding layer which is a 5-vector and out inputs which are 4-vectors and we’re trying to pull out, so it is a genre aware attention model so we’re taking our genre which I guess is a 5-vector.

Interviewer: I know there isn’t enough here, but would you say that is a 5-vector.

Participant: Yes I would say that is a 5-vector and the inputs are all 4-vectors.

Interviewer: It’s represented h1 hk and hn. Would you get the same information if it was just hk? If it just represented just the kth element.

Participant: Umm, objectively yes, but I prefer having the multiple. And actually no, I wouldn’t because the sum element wouldn’t have translated. You could probably have worked it out, but it would be a lot less obvious.

Interviewer: Next one?

Participant: Ok [Laugh]. [B Pause] I like the figure description. I suppose I should say the first thing I did here was to read the description before trying to figure out what was going on. [Pause] Ok what is being shared? [Pause] Umm so ok first impressions, I’m finding it vaguely tricky to match the description to the diagram. The two fully connected networks on the top left, fine. The bit that is in black and white [Pause] not immediately obvious. It seems to indicate those two in the black and white, the right hand half are shared, and the same convolutional layers. But it’s not obvious why. That should be 1st person views on the left? Maybe? Two things are labelled third person view, so where is the first person view. Ok so we’ve got two streams. [Pause]. Ok, to be honest I reckon, if my interpretation is correct and the right hand part is actually shared, then I don’t like that representation at all, why don’t you just draw it once to represent the fact that both of the signals are passing through it. The embedding space, yeah that’s quite cool. Although the arrows between are a rubbish way to show the difference between two vectors. There are much better ways. Fine. Ok so you’ve got two video streams and you have two networks and they share some layers and you take two different points in your embedding space and there you go. I feel like there is something I don’t know about. Some of the techniques here for identification that would maybe fill in the gaps if I knew about them. The grey lines between the green and the yellow blocks. [Pause] Yeah it’s not describing the siamese structure. I don’t know what that is doing, so that is not helpful. [Pause] Ok so do I like it? Yeah it is nice and linear. It’s nice and linear, but has not revealed much more to me than the text caption did to be honest.

Interviewer: So things like underneath each cuboid there is a number.

Participant: The size of the filter I assume. The size of the convolution layer I assume.

Interviewer: Is that useful? Is that the kind of information you would like to have in this diagram?

Participant: Yeah. I suppose it helps understanding because it is what I would expect to see in a convnet diagram. So it helps me confirm in my head that those are indeed convolution layers.

Interviewer: So does it matter what the numbers are? Or is it that it complies with an expected convention?

Participant: The numbers don’t matter to me at all. No. Pretty irrelevant I would say. The ratios are fairly important. Or at least the ratios are fairly interesting. The exact numbers, the absolute numbers, no.

Interviewer: Have you seen this type of diagram before? You already said it was conventional?

Participant: Yes. The convolution net diagram.

Participant: Ok C. [C Pause] Wow. That’s exciting, lots of stuff going on there. Ok, read the text. [Pause]. Cool, alright. [Pause]. Ok I’ve never come across the two triangles pointed at each other, but it’s been labeled, so that’s ok. [Pause]. Slightly confusing there are two things called Dg, if I had the paper I would read a little bit to try to understand why there are two different things called Dg, two different things called Df. Iteration from cloud fusion. Ok. Don’t know where these clouds are coming from. [Pause]. It’s kind of neat how they’ve picked out the distortion. I guess from that I can imply there is some other stuff going on that’s been described somewhere else in the paper, that I would probably go and glance at. [Pause]. Ok cool. [Pause]. So I’m getting the general gist of what is going on here, don’t know about all of the details. Presumably, I’m assuming here, that the dotted lines at the bottom are saying “for sequential images” or something. I don’t see why that would make sense in this case. [Pause] The colouring of the guided disparity map, unless they’ve used other images in the same paper, it would be helpful to have a legend. I’m assuming that they have a car which is highlighted in pink is the most interesting thing to look at. [Pause] Anything else I can say? [Pause] I don’t quite see how those particular stereo images relate to the clouds. If they do at all. [Pause]. Yeah not much more to say there, I think.

Interviewer: What do you think of the use of an example in this one?

Participant: Yeah, I like it. Although like I say I don’t quite see how the cloud maps into that particular picture. Like it is saying projection of what onto what. We’re taking a bunch of clouds, and there are some examples, fine. I can see how those clouds are being fused into a multitailed cloud, fine, whatever. But I don’t see how that cloud is being projected onto the guided disparity map Dg. And I don’t see how, we’ve got Dc and Dg, I don’t see how we’re going from iteration 0 to iteration 1. Primary and secondary filtering. [Pause] Ok so Dc is without any filtering. I have no idea what the filtering is doing, but I guess the fused cloud is being applied somehow.

Interviewer: Shall we move on to the next one?

Participant: Sure. Sorry one other point. The yellow line connecting Df back to the fine tuned guidenet, no idea what that is doing there, and why the other guidenet doesn’t have one, but yeah.Otherwise that’s about it. D.

Participant: [D Pause] Oh that’s exciting. [Pause]. Ok, quite cool. [Pause]. The legend is way too close to the diagram, took a few seconds to realise what that was doing. [Pause]. Cool. [Pause]. Now those look like Haskell-type signatures above the decoding layers, which would make sense. [Pause]. It’s a bit odd that it is going bottom to top. [Pause]. Quite cool. [Pause]. I wonder if those two, the memory schema switch, are connected. I can sort of see what it’s doing here, but I wonder how it is switching between memory and schema. “As shown in the figure”, really. [Pause]. Ok, so I quite like it. There’s a clear use case and example question, so I can figure out what problem it is trying to solve and I can sort of understand how it is solving the problem. Some embedding layer, which is these grey blocks, and then there is an encoding layer, which is these yellow things. I don’t quite see what the arrows between the encoder units are doing, but ok. Then we get some vector of the question encoding, and then it is going through this pipeline of transformations which is pulling out the interesting elements and then applying something. Doing some thing, applying some filters, selecting some columns. [Pause]. Like I said, the memory versus schema, I can’t quite tell if that’s a temporal thing or what is controlling that switch. The little plots below the columns, showing the embedding match, is pretty nice. I can only really make sense of it because I kind of had to look for that. I’m assuming those grey and blue columns are some measure of how well the column matches the embedding. And the same for the memory. So measuring how well that element of the text matches that thing. Slightly misleading because it sort of makes you think, I initially though the decoder units mapped one to one to the question elements, they don’t of course. That was slightly misleading, because the embedding and encoder units do. That’s probably about all of the useful information I can pull out of that. [Pause] Also why was book club not in the memory schema.

Interviewer: One thing I’d like to pick up on is the bottom to top. What is your expectation around orientation or flow in these diagrams?

Participant: Normally top to bottom, left to right. That’s what I’d expect.

Interviewer: So here starting on the bottom right or bottom left?

Participant: It kind of is bottom right. My head is going bottom right, then around the corner, then back across out. I kind of had to get my head around it, because it was not what I was expecting.

Interviewer: And so overall you like this one?

Participant: I quite like it. There are still some gaps, but I get what it’s doing and I can kind of see how it has been implemented.

Interviewer: Next one?

Participant: E. [E Pause] Hey, there is a main structure of our model. Great [sarcastic]. Cool, so we have lots of stuff going on here. Many different things being used. First thing I’m latching onto is top left and bottom right, which is great, I can see what we’re taking in and what we’re taking out, which is great. [Pause]. Got resnet, and some BERT. Wow, lots of stuff going on. Big feature matrix. Why are there three colours? [Pause]. Ok. [Pause]. What can I say about this one? There is a lot of stuff, and I think they actually do a pretty good job of explaining what is going on at a high level. There is a lot of detail that I have completely no idea what is going on, but could make guesses at. Like the procedure captioning module, some bit of a feature matrix going into an encoder that is also taking in a bit of a transcript, and straight into a decoder. I’m quite interested to know what is going on there. Probably irrelevant for the overall thing, but that seems quite interesting. And the transcript goes back to there. [Pause]. The fusion module, meh. Vaguely can see that they are combining the two resources somehow. Having all these 10 rows, or whatever that is, doesn’t add much important really. It tells me there is lots of data, but don’t know what that means. Feature matrix, ok so different parts of the feature matrix are being mapped to different sections of that embedding space. [Pause]. What’s happening? [Pause]. Hmm. Ok so there is an LSTM, great. I can’t quite figure out what the inputs to that LSTM look like. Is it low resolution? I don’t think so, zooming in there. There two different textures on the cells of the inputs to the LSTM, I’m not sure what the difference between spotty and stripey. [Pause] Length. [Pause]. Ok back to big picture, what is this showing? There’s some, probably the most useful bit I think is the t-start to t-end, taking a chunk of the feature matrix and a chunk of the matrix, encoding and decoding out to the procedure caption. At a specific time, I assume. Yeah. I suppose overall, I can vaguely tell what is going on, but I have hardly any idea about the details. Could they have described this better in text? Probably not. It is nice seeing the different modules. [Pause]. That’s kind of good. But I’m not clear on how they join together.

Interviewer: When you say you like the modules, you like that they have boundaries?

Participant: I suppose I like how they’ve been broken out into sensible, separate chunks.

Interviewer: The level of detail in here, is it appropriate or helpful?

Participant: Some of it is helpful and some of it is not. The context aware fusion module, I mean those rows don’t add anything that I can tell, so that is not particularly useful. I think you could replace that with just a block. The level of detail. It is useful that they are describing, I vaguely get inputs and outputs, but again not everywhere. It’s kind of a mix, in some places there is good detail, and in other places there is no detail at all. Like transcript embedding, there is a relative amount of detail in that, whereas the BERT model. Less detail would just be a block saying BERT, with an implicit 3-vector output, whatever. But I think that is quite useful to see there are lots of arrows there. And it conveys the sense that there is a lot of data being moved around.

Interviewer: So overall?

Participant: Overall do I like this picture? [Pause]. General feeling is yes, but it comes with a caveat that it doesn’t really tell me how to implement it at all, or how I’d replicate this. It’s just at quite a high level a decent picture.

Interviewer: Final one?

Participant: [F Pause] It would be nice to have a bit more detail about what merge is. [Pause]. Fine. Pretty standard I suppose. Yeah. [Pause]. I like it because it is simple but that doesn’t necessarily say anything about the image and more about what it is doing. [Pause]. Yeah, so could definitely be made more pretty, but it tells me what’s going on in an ok-ish level of detail. If I saw this in a paper, it would give me a good indication about whether I wanted to read the text and figure out what is going on, because I can tell, I can understand in enough detail what is going on. The interesting thing, the stuff I would expect, the FC layer at the end and softmax, fine. The merging layer as well, but it would be nice to have a bit more detail there. I don’t know if they just concatenate or what. [Pause]. Yeah ok I mean fine, there’s two input streams. It doesn’t tell you much but that’s probably a good thing.

Interviewer: So you mentioned about aesthetics. How important are aesthetics for you in diagrams in general?

Participant: [Laugh]. I was going to give a cop-out answer. They’re really important if they impede my understanding. Total cop-out. No, I think they are. Personally I appreciate a well designed diagram. For example, it slightly bugs me that the typeface is slightly smaller in projection layer than bidirectional LSTM. I mean that’s not hard to fix, you could’ve spent an extra two minutes to make that text bigger, but that’s just being picky. In this case, nothing here impedes my understanding from a process point of view. I can see what’s going on. They could’ve made it prettier but it wouldn’t have changed anything about the information conveyed to me, I think.

Interviewer: Merge is in an oval, and softmax has a dashed line around it. Are those things significant?

Participant: Beats me, no idea. Presumably they are significant, but yeah I suppose they didn’t jump out to me as things I would want to find out more about. Except for the fact that merge is totally ambiguous. I guess softmax I ignored as it is relatively implicit.

Interviewer: If the merge had instead been concatenate, what would you think then? What’s the next thing you want to know?

Participant: [Pause] I suppose [Pause] I reckon the next thing I’d want to know is how big are these things. Actually, next thing would be can I have an example. At some point I’d want to know how big are these, how big are the vectors. And how big is the FC layer at the end. Yeah, examples and sizes. I’m not particularly interested in the architecture of the LSTM. Because they haven’t given more detail, I would assume that it is just a standard BiLSTM. [Pause]. Projection layer might be good to have a bit of detail there, but I’d just assume it is standard. Mean over time, at some point I’d be wanting to ask what does that mean. What does time mean there. What does the mean over time actually mean. Are you averaging over some number of words or sentences or something, or what?

Interviewer: And so overall?

Participant: Yeah, I like it.

Interviewer: The next part is a tasks list. If you could pick out the top 2 or 3 things that are important in your use of diagrams, and anything you definitely don’t do. If there is anything missing from this list, particularly as you are a different sort of user to most, we can definitely add that.

Participant: [Pause] Ok so this is both things I use in my diagrams and things I look for?

Interviewer: When you’re using them.

Participant: Ok cool. [Pause] Right so identifying how it works and the purpose are near the top. [Pause]. Identifying system novelty, which isn’t always straightforward. It is not always obvious, because of what you’re doing this research for. I would still put that near the top. Top 2 or 3 are definitely around the purpose and features. The least important, there are a few. I don’t use as index to navigate the paper, can’t say I’ve ever done that. Author thinks, never really thought about it in those terms. Corpora and data types, not useful. And the rest is probably somewhere in the middle. Is there anything else, I’ll just have a quick think. I’ll sometimes glance at the images to get a feel for the overall complexity, which I guess is part of how it works. I’ll look at the title, or I’ll read the title of the paper and think it sounds very interesting, and then I’ll have a look at the diagrams to see how visually complex it is. Based on that, I’ll make a judgement about whether I think this is complete nonsense, is this a paper I want to spend time reading, or have they done something completely obvious. So I suppose if the paper is worth reading, the diagrams will usually have something that is noticeably different from other papers or the usual. If it is just a diagram. This is all about how the system works, it’s all about that. If the diagram is what you would expect, from the title, cool, I’m down with that.

Interviewer: This is interesting because it is a subset of the how. The overall complexity you’re after. If you saw something very complicated, like the 5th one, would you say that it isn’t even worth it?

Participant: If the title of the paper was something trivial, and they came out with a crazy complicated diagram, I would probably be a little put off, because it would seem like they are overcomplicating something. I suppose I would compare it with what I was expecting from the paper. Complexity is fine, complexity in a diagram is totally fine, but [Pause] I suppose if it is complexity without a reason, then I am generally turned off. If the diagram appears to be overly complex for no particular reason, then I guess that puts me off the whole paper.

Interviewer: Any other things about your usage of diagrams, in consuming diagrams?

Participant: Yeah, examples are great, especially for image processing. I can think of one in particular where I didn’t get the paper, and I was reading it and got to the examples near the end, where it walked through the network and the outputs of each stage, and it suddenly made sense. I suppose that is one where it was definitely great to use a diagram for that.

Interviewer: There’s a last couple of catch-all questions, which you’ve already alluded to. Do you use diagrams to reason as part of your creative process when doing research? How?

Participant: Yes. If I’m trying to understand how something works, or come up with a way of doing something, I’ll sketch it out. It’s very helpful to have a visual representation for flows. Computation in general. So I’ll definitely do that. Also with machine learning I do find that if I’ve got some new concept to try to reason about, having a defined way of diagramming is really helpful. If there is a particular type of network or particular type of graph that I’m using to describe a particular application, then having defined elements for those diagrams is helpful. Then I can look at ten of those diagrams and know that they are all describing the same kind of thing and all using the same conventions. I do enforce that when I’m doing my own notebook diagramming, I’ll use a consistent set of conventions for myself.

Interviewer: That’s just for you, not the wider organisation?

Participant: Generally it starts off as just me, and then becomes part of the wider organisation when I share stuff. Some of it will be taken from well known diagrams, conventionally used diagrams that everyone knows. There’s nothing novel there, the main thing is it is self-consistent.

Interviewer: The self-consistency is useful because?

Participant: Because I recognise all my diagrams and I don’t have to think about what they mean.

Interviewer: Is it to compare them?

Participant: Is it to compare them? Yes to some extent. It is also to make them easier to draw, because I stop looking at what the shapes are and start thinking more about the meaning. It’s useful to compare them, definitely. It’s useful to reason about them as well, just because I suppose the abstraction just melts away, and you can understand what’s going on rather than the shapes.

Interviewer: So do you think this is a problem, without being leading about it, in the existing literature?

Participant: That there isn’t a more consistent set of things? I guess one of the first things when I open a new paper, and this is a problem with mathematics in general as well, the lack of consistency in notation, when authors decide to arbitrarily adopt their own notation for things. The first thing is often to understand what symbols and shapes mean in a paper. It would definitely be helpful if there was a more consistent format for all of these things. I suppose like the way LaTeX made typesetting recognisable and standardised, you know where to look for certain things.

Interviewer: Final one is can you think of any tools that would help you in making or consuming diagrams?

Participant: Computer tools?

Interviewer: Tools in the broadest sense.

Participant: If I could sketch stuff in my notebook and pull that into my computer straight away, that would be helpful, and continue editing it on my computer, that would be helpful. For producing. And like I say the programmatic generation of diagrams is useful, but some tools do already exist for that. More would be good. If there was a standard way of describing network diagrams, maybe there is and I don’t know about it, that would be useful. The problem there that is inevitably arising is that you have some kind of block that is novel, a new architecture, and you need to integrate that into your diagram. So I suppose tools that would make that easy would be great. Consumption, [Pause] I can’t really think of anything around consumption, no. Generally just look at the diagram.

Interviewer: Anything else you’d like to say before we finish?

Participant: No.

# Participant 11 (P11)

Interviewer: How would you describe your experience level with Computer Vision and Natural Language Processing?

Participant: Pretty high, I mean my job is in Computer Vision, so that’s what I do for a living. Obviously I have quite a lot of experience in that. Natural language processing I’ve done a few online tutorials but not done in anger. Computer Vision good experience, NLP not so much.

Interviewer: And you’re a postdoc?

Participant: An RA, yes.

Interviewer: How do you use diagrams when communicating your research?

Participant: Probably not as much as some people, I tend to be quite verbose, as you know. I prefer to explain things but there’s no question that there are some things that a diagram can just explain really succinctly and clearly. I would probably err towards communicating things verbally, though I absolutely see the benefit of a good diagram. As a friend of mine, who was very influential during my PhD says, sometimes a really good diagram can make a paper, or break it.

Interviewer: So when you’re including a diagram in your papers, does that happen most of the time or some of the time?

Participant: I’ll be completely honest, I’ve not published any papers since my PhD, that is actually about to change but there were diagrams in the papers I have published and yes those were important. I remember I had to do a poster for a conference, similar sort of thing, especially in the poster, people aren’t going to read blocks of text, they want to see diagrams, so yes I’d say they are very important.

Interviewer: What are these diagrams of?

Participant: Well there’s the obvious ones like diagrams presenting the results, but also just a diagram to describe the network that I’ve built, the neural network. For a while there were various different ways you could do that, and there’s a sort of unified approach to it now, with various styles of blocks. Because there is a unified sort of approach to describing it you can look at a diagram and see straight away what you are looking at, so that is very helpful.

Interviewer: So that’s for image processing system diagrams?

Participant: Yes, the diagram I’m talking about there would be the diagram to describe essentially the model parameters, the network parameters, the number of filters in the layers, that sort of thing.

Interviewer: For like a CNN, where it’s like 3D blocks?

Participant: Yes, that’s exactly what I’m thinking of.

Interviewer: When you’re making one of these system diagrams, what’s included in that, if you’re drawing a diagram of the system?

Participant: [Pause] Whatever is required to get the parameters of the system across. As I said, I’ve not published in the last couple of years, so I’ve not had to do this. Whatever diagram I was putting together, the main thing I’d be looking to do is to convey what forms that model as succinctly as possible. Now as I say you’ve got this unified approach to neural nets, that’s quite simple now. The models I’m working with now are using random forests, so it is not quite so easy to draw a diagram of those. We have a reading group every week, and when you see one of the diagrams that we’ve been talking about, you get straight away what’s being discussed.

Interviewer: So what sort of problems do you encounter when you’re making these diagrams?

Participant: One of the problems is the number of layers you’ve got in neural nets these days. You literally can’t include them all in a diagram because they’ve got so deep. So that would be an issue. As I say there’s not a huge amount I can answer there because I’ve not done too many diagrams.

Interviewer: Would you include input and output as part of every diagram, or if it isn’t important would you omit it?

Participant: I think that is important. I certainly would include input and output, they’re very important.

Interviewer: And what about the level of granularity?

Participant: I think that is very much dependent on how big the network is itself, and how much space you have available. If you try to put too much granularity and have a very complex network. The most important thing is that the diagram is clear. If you put too much granularity in there, it won’t be clear, which ruins the whole point of the diagram.

Interviewer: So when you’re making diagrams, clarity is the important thing?

Participant: Absolutely. If a diagram is not clear and obvious then that wrecks the whole point of the diagram. Diagrams are supposed to make things clear.

Interviewer: Do you think that diagrams should stand alone in a paper, or should it have reference to the text as well? Should you be able to understand everything in a diagram?

Participant: I think that depends on the person who is writing it, and the paper, and also the material they are trying to represent. [Pause]. Let’s say for example the standard NN diagrams we’re talking about, those probably can be a lot more stand-alone now. But it might be useful to tie you into the text more if the diagram is helping you. I don’t think there’s a rule, the answer is you can use for what you require.

Interviewer: When you’re making diagrams, what sorts of tools do you use?

Participant: Really fiddly ones. I have used MS Word, the Paint tools in that, but they are notoriously horrible. I’ve used inkscape, that’s probably the main one I’ve used.

Interviewer: And why do you use that?

Participant: [Pause] With the MS Word Paint tools, that’s when putting a very simple diagram together. When I’ve had to do something more sophisticated, something for print, certainly something going in a LaTeX document or something like that, that’s when I’ve used inkscape. It isn’t a particularly nice piece of software, I’m sure there are better packages out there, but that’s what I’ve used.

Interviewer: So what about when consuming research, how do you use diagrams?

Participant: I find them really helpful. I struggle a little bit with text. I don’t know what it is, if it is the font or what, but I actually find it quite hard to read research papers. What I tend to do now is put them on my laptop and dull the screen a bit. But anyway. I really like a diagram in there. If you’ve got a good diagram in there that conveys that information, I find it much easier than reading the text. I think it also because academic writing is so dry, and I don’t enjoy reading that. The first half paper I wrote, I was actually told perhaps don’t use sarcasm in scientific papers, I disagree [Laugh].

Interviewer: That you enjoy reading papers with diagrams in it, and that you enjoy the visual elements, seems to almost stand against your previous comment about liking to be very verbose in your own work?

Participant: True actually. When I was talking about communicating my own work, I was more talking about person to person communication as opposed to written. Actually it doesn’t, my ideal way of communicating my work is verbally, whereas if I was to do it in writing I would prefer to use diagrams. I wouldn’t rely on diagrams, my verbosity is always going to make its way into that paper, but I think diagrams are very useful in that regard. I was more talking communication person to person.

Interviewer: So if you had to say what kind of learning style you had, like audio or visual or kinaesthetic, how would you identify yourself?

Participant: I learn by just getting on with it, I’m kinaesthetic. I just throw myself into tutorials. When something in a tutorial doesn’t make sense, I’d maybe find a youtube video or something like that.

Interviewer: How do you read a paper, could you talk me through what your process is?

Participant: [Pause] I tend not to bother with the abstract, I tend to just get stuck into it. I don’t tend to read the background a huge amount, I’ll flick through it but I’m more interested in getting straight to the methods section and seeing what they’ve done. I’ll probably start with the introduction, because obviously that gives you an idea of the paper. I will look over the background section but it’s more, very often it is talking about work I have no idea about, and unless I’m going to go down various rabbit-holes to understand those works then there’s, you know. I want to know what they’ve done, and as time passes I’m spending more time on the results section that I used to. I used to look at the results section and if one number was higher or lower I’d be like “oh that’s great” but now I’ll be like “oh that’s actually within margin of error on that” and start to, as my knowledge of how numbers work improves, I’m spending more time looking at what the numbers are saying, if that makes sense.

Interviewer: And if you come across the diagram in the process, then you’ll look at the diagram?

Participant: Yes. I find diagrams in the methods section particularly useful.

Interviewer: But it’s not like you’re using the diagram to pick out straight away, looking to see if there is a picture?

Participant: No.

Interviewer: So the diagram becomes part of the text in some sense?

Participant: Yes.

Interviewer: What other problems are there in consuming research more generally? I guess you’ve already spoken about there being a lot of stuff, but if there are other topics you think are interesting?

Participant: I probably should spend more time on this, [redacted for anonymity], the literature review. I love to get involved with the coding and just get stuck in, and I probably should do more of checking the literature, seeing what else has come out, I don’t do enough of that.

Interviewer: Shall we move on to the diagram series? There are some example pictures, tell me what you like and don’t like?

Participant: A. Well this is actually quite similar to the style I used in what I did before they came out with these very standardised ones. [Pause]. I mean with this one you can see. [Pause]. It is pretty clear, an older style of doing it. I’m used to seeing these, others might struggle with it if they weren’t used to seeing this style. The only thing I don’t like with these, and I didn’t like it with my diagram, is where you’ve got these circles and rectangles. I’m working on the assumption that there aren’t four inputs there, and it’s kind of an arbitrary number. Whereas with the more blocky ones you can put 256 and that gives you more information. You’ve got information on that diagram that I’m guessing doesn’t actually mean anything. Like the output here you’ve got nine outputs going to the sigmoid, and I’m guessing it isn’t just nine. Maybe it is. And that’s where I think these fall down a little bit.

Interviewer: So what is it that you like about it?

Participant: You can see the flow of information from bottom to top, you can see the different layers and how data is flowing from the start to the finish, that is pretty clear on this.

Interviewer: And the labels on the side?

Participant: As I don’t know what Genre aware attention means, I guess that would probably make more sense if I had the paper. Input dense layer. I can guess what the dense layer is, but that isn’t just going to stand on its own, I would need more information than that.

Interviewer: What about the use of mathematical symbols?

Participant: Yeah, happy with that. Like it.

Interviewer: Anything else that you don’t like?

Participant: I don’t really understand what the dots [ellipsis] mean. That means there’s lots of. Ok that does actually make sense. It’s ok.

Interviewer: Do you like the use of colour in this?

Participant: Yes I do, I like a bit of colour in things.

Interviewer: Next one?

Participant: B. This is the more standard style of NN diagrams. [Pause]. Yeah this gives you the details of. It’s alright, but it’s not quite all the information. You can tell how many features it is picking out in each layer, but it doesn’t give you any idea of size of the patches or anything like that. You can see what’s going on with it definitely. You’ve got an input, it is moving through various layers. Again it doesn’t give any details of whether you have pooling or anything like that. It is using the sort of style I would expect, but there is information on there that isn’t on this one. So, it’s ok.

Interviewer: So the information you’re expecting to see, is it important information or is it just strange that it isn’t there?

Participant: I’d say it is quite important, because it gives you an idea of what is going on in the network. As you move in pooling layers, as opposed to convolutional layers they’re very different, and if you have no indications of whether these are in evidence on this, so I’d say it’s important yeah.

Interviewer: It feels like you’ve seen this before a lot, do you think that familiarity is helping you?

Participant: Yes. This is what I was talking about as a unified approach to doing these diagrams. In actual fact, on my PhD thesis one of the corrections they asked me to do was to redo the diagrams in this style. I understood why they were asking that.

Interviewer: And what’s your thoughts on use of colour in this diagram?

Participant: I don’t know if it really adds anything. Why are these white then become yellow? The colour seems a bit arbitrary. I do like a bit of colour, but there are colour changes. I think what it means is that this is some kind of Siamese joint network, so it is trying to differentiate between those bits and those bits. You’ve already got something which indicates the network is changing at that point. I think they maybe overuse colour in this one.

Interviewer: What about the contrastive loss visualisation?

Participant: That is pretty awful, to be fair. It looks like they spent loads of time on the rest of the diagram and were maybe in a bit of a hurry to throw it all together at the end. That’s what that looks like.

Interviewer: And the shared arrow in the middle?

Participant:I can guess what that means, these is information being shared between the two sides of the Siamese network, so yeah that’s ok.

Interviewer: Overall what’s your feeling?

Participant: It’s ok. It is using the standards of laying out a NN, but I don’t think it is a particularly good example of laying out that standard method.

Interviewer: Next one?

Participant: C. This is a bit of a colour crash. [Pause]. Doesn’t help that as I’m getting older I need to get new glasses. It’s ok, it just looks a bit all over the place. I think there is a bit too much colour in this one. [Pause]. Without the paper, I mean I can guess what’s going on but I’m not massively fond on that one.

Interviewer: What’s your thoughts on the use of an example throughout?

Participant: I think that actually nails on the head what I don’t like. It may be better to have the example separately. It is one thing. An image should either describe how the model is constructed, the nature of the model. And then you can have the details of the images and their processing can be done elsewhere, I think you’re trying to put too much information into a single diagram there.

Interviewer: Is there anything you like about it?

Participant: [Pause] It’s alright. It’s not the style of diagram I would use. Not really. On the bottom right, I’m certainly struggling to see what the difference between those two is, condensing to a smaller level. You’d really have to zoom in I think on a computer to really get any benefit out of those. On a piece of paper it quite hard to make them out.

Interviewer: So nothing?

Participant: Not really, no. It looks a bit of a mess to me.

Interviewer: Next one? D.

Participant: There’s too much information on this one. [Pause]. This here, the embedding and encoder decoder units, if that is a key to show what the different ones are, they really needed to move that away from it. Initially that looks like it is part of the network. [Pause]. Yeah this is another one where if that diagram is in a journal or a paper, you’re really going to have to zoom into that, it would be unreadable in a normal size. There’s far too much information. The whole point is it is meant to be concise and get the information over to you quickly but that it not concise at all. [Pause] And it’s quite hard to tell what it is actually doing. This is the sort of diagram you’d need the text to accompany it to understand what is going on there.

Interviewer: Is there anything you like about it?

Participant: Not really. It just looks a bit of a mess. It is quite hard to make out. I don’t like it when you’ve got things moving over. Flow going left to right to left to right. For me, flow should go one way. So either left to right or bottom to top. I know in different cultures they do right to left or whatever. But it should just be one. If you start going up down and all around, it looks like a big mess to me. [Pause] Yeah, it’s quite hard to tell exactly where the output is. Yeah, I’m not a fan of that diagram.

Interviewer: Next one? E.

Participant: It’s ok. Again, I think they’ve gone into too much. As opposed to going into lots of boxes and things like that, you can have a single box and put the numbers next to it. I think that is a much better way of conveying the information. I think half of the problem with a lot of these diagrams are they’ve, and it is something I’ve been guilty of in the past, is they’re putting too much information into a picture. You want to get your point across, as I’ve said a couple of times, succinctly. So for example these, you have ten by two circles in the CAF module. Now, I mean if it is 10 by 2 then it means something, but I’m guessing it probably doesn’t. You could make it 5 by 2. You’re putting information in there that is unnecessary and cluttering it. Again under procedure extraction module, you’ve got all these boxes, and I’m not really sure what they represent. Is that 7 by 6, is there any importance in the blue 7 by 6 and the lower being 4 by 6, or is that arbitrary choice being made? I’m guessing it is arbitrary choice, in which case you’re putting too much information in there. That would be better with 3 rectangles with the numbers next to them giving you the actual dimensions, that’s what I’d say about that.

Interviewer: And the use of colour in this?

Participant: The use of colour is ok. They’ve not tried too much. It’s all pastel colours, so you’ve not gone for the full “throw all the colours on it” and that works better I think.

Interviewer: What about in the transcript embedding, being 1 and n and skipping out the ones in the middle. Is that?

Participant: That’s fairly clear yes. The fact they’ve used 1 and n does signify, does imply, there is stuff in the middle as well. So yeah, I initially slagged it off, but the use of dots in the previous one might help. When you’re doing a mathematical equation 1, dotdotdot, n that’d work quite nicely, and I guess that is what they’re doing here.

Interviewer: The input and output example?

Participant: That’s ok. That’s actually the first thing I probably looked at, which tells me they must’ve done a half decent job of getting across how the information travels through the network, so yeah, that’s ok.

Interviewer: This macroscopic module structure, does that work for you, in terms of reducing complexity?

Participant: Yeah, I mean the dashed rectangles round them does give a sense there are different parts to this model and that does work ok. I just think it is a bit cluttered, because there is a lot of information that probably isn’t necessary on there.

Interviewer: In terms of the actual content, like for this pretrained BERT, offset length, etc, is that information that you want, and is there semantic content you feel is not included?

Participant: I don’t know enough about this network to know. I’ve not used resnets so I don’t know what resnet-34 means, and I don’t know what a pretrained BERT is.

Interviewer: Fair enough. Anything else about this?

Participant: Not particularly.

Interviewer: Next one. F?

Participant: That just looks a bit lazy. [Pause]. Yeah that is not giving much information at all. The fact that both top and bottom layer are exactly the same, the word projection layer in both, that just comes across as quite a lazy diagram. One thing is bidirectional LSTM, looks like the information is only going left to right. That’s a crap diagram, as far as I’m concerned.

Interviewer: You’re saying it’s crap because it is too [Pause].

Participant: It’s not really giving you [Pause]. It’s not giving you any extra information that you couldn’t get in a sentence. You could just say “this uses a projection layer, a bidirectional LSTM, mean over time” I don’t know what that is. That could actually be described. The diagram isn’t giving you over and above what you could get from text. So in the sense of conveying information clearly, because it is that simple and basic it isn’t giving you anything a simple sentence wouldn’t give you.

Interviewer: Overall, block diagrams that are like this but have more information, would that be ok, or is there [interruption]

Participant: Definitely yes block diagrams are fine, it is just there isn’t much going on there. At the top you’ve got the words “fully connected layer” and exactly the same words underneath. “Merging layer” and “merge”. It’s just really basic. You didn’t need to have the top bits for a start, you can get rid of those, it’s just repeating information. Bidirectional LSTM, it just [Pause] looks very [Pause] there’s not much information being conveyed in that.

Interviewer: Is it the specificity of the information that you want?

Participant: Yeah, I’d like to see more information about it. Bidirectional LSTM, does that really give you much information about the nature of that LSTM? There’s no information about parameters there whatsoever, it’s just blocks, no numbers.

Interviewer: So you’d like parameter and hyperparameters?

Participant: Yeah, in the way that the CNN diagrams have the size of the features, whenever it is looking at things like that that is quite useful information.

Interviewer: Great. Last part is task sorting. If you can rank these in how important they are in your use of research paper diagrams. Focus on the top 2 or 3, and any you don’t do.

Participant: [Pause for exercise] Ok so I’d say those three. Identifying specific architectural features. Understanding how the system works and identifying the layers between components and internal dependencies.

Interviewer: Any of these you don’t do?

Participant: Index to navigate. No that’s ok, ignore that. If you see a diagram of a paper you know that you’re in the method section, so that is probably useful. [Pause] I mean all of these could potentially be useful. Those three are the top.

Interviewer: And all of the others you do sometimes?

Participant: Yeah I mean I don’t know if I’d use the diagram to write a summary of a paper. System novelty, if the diagram helps with that, then great. I’ll stick memory aid at the top, as that’s useful. I tend to remember things more by pattern, and diagrams can be quite useful there.

Interviewer: Are any in the do-not-do pile?

Participant: [Pause] There’s an argument they could all be useful. You wouldn’t do all of them because then you’d be trying to put too much information in the diagram. Depending on what you’re trying to get across, each of these could potentially be uses. I wouldn’t rule any of them out, those four are the more important.

Interviewer: Is there anything else that you use diagrams for that hasn’t been captured here?

Participant: Not really. The main thing for me is specific architectural features and layers and internal dependencies. That’s pretty much covered it.

Interviewer: The last part is some catch all questions. Do you use diagrams to reason as part of your creative process?

Participant: No.

Interviewer: No sketches?

Participant: No, for me I do this very much inside my own head. I am quite good at visualising stuff myself, so I don’t. I tend to do it in my head.

Interviewer: Do you use diagrams in any other way during your research more broadly? You already spoke about posters.

Participant: Not over and above what we’ve discussed.

Interviewer: And can you think of any tools that would help you in making or consuming diagrams?

Participant: I suppose it would be quite nice if we had an improvement on inkscape. I suppose maybe someone has and I’m just not aware of it. A piece of software which does the NN diagrams much easier, that would be great. But, because I’ve not had to do one for a while, if I was to have to do one that would probably be the first thing I would do. But beyond that, not really no.

Interviewer: So if you were looking to make a diagram today, how would you go about that?

Participant: First thing is what information do I want to get across, and what do I think will be most useful to the reader. Then I’ll just figure out the best way of getting that information across. Getting that balance between getting the information on there but not making it too cluttered, that’s hard, I’ve not had a huge amount of experience doing this in anger. One of the first things I’d do is get some advice from someone who is more used to it than I am.

Interviewer: And would you be looking at what other people had done?

Participant: Yes, definitely.

Interviewer: Anything else about diagrams?

Participant: Nope.

# Participant 12 (P12)

Interviewer: The first question is a bit of a background question. How would you describe your experience level with computer vision and NLP?

Participant: So I would say with CV I have published one paper on the topic, but I don’t have much more experience than that. I have been toying around with things, but not really much in detail at the research level, I would say. I feel confident that if needed I would be able to train a classifier that was using neural networks and I’m doing that, but not much more to be honest. And also I work with GANs, but that’s what I did my degree on, so I have confidence on the topic. On the other hand, for NLP, I’ve done all my PhD on the topic, and all my research career on the topic, with several papers, so on that I feel much more confident, let’s say.

Interviewer: So how many years experience do you have in NLP?

Participant: If I had to put numbers on it, I’d say 10 years NLP and 1 year on CV.

Interviewer: So, substantially stronger in NLP than CV, but with some broad base?

Participant: Yes, I’d say for general ML, NNs and deep learning, the parts that are common among these fields, I am confident on all of them.

Interviewer: Your job at the moment is as a researcher?

Participant: A research scientist, yeah.

Interviewer: And your job at the moment involves reading and publishing papers?

Participant: Yes, definitely.

Interviewer: Do you build software as part of that?

Participant: Yeah. I write prototypes for paper stuff to begin with, and also an open source tool called [redacted]. [redacted].

Interviewer: I read this paper, so I’m looking forward to talking to you about it a little bit. We could even talk about this a little bit now. It is how you use diagrams as part of communicating your research. I noticed in the [redacted] paper there are some block diagrams that make use of colour, and it would be nice to hear how you went through that process.

Participant: Yeah yeah yeah, so let me think about it. Maybe there are three recent papers we can talk about. [redacted]. Regarding the [redacted] paper, there is one thing I do in [redacted] that is trying to organise. Basically [redacted] has a general structure which is supposed to match many different NN architectures. The idea is to increase the level of abstraction so that different neural architectures are interchangeable between each other. And so the diagrams in that paper basically describe this higher level of abstraction which I define as [redacted]. [redacted]. So the diagrams describe this higher level of abstraction and not each of the layers of the model.

Interviewer: So why did you choose to use a diagram rather than pseudocode or an API language?

Participant: I actually use both. In the description of both of those in diagrams in the rest of the text of the paper, I try to do that, to specify the inputs and outputs, and what the inputs are and what the outputs are for each of those models. But specifically for me there is the second image of the paper. There are two main images of the paper [redacted], the first describing the general architecture. I think the strength is in the second image, where I show that combinations of different things map to different ML problems, with text input and class output is a text classifier, with image input and text output is an image captioning system, and so on. So the diagrams visually show that in a much more visually powerful way than text would do, I guess.

Interviewer: “In a visually powerful way” what do you mean by that?

Participant: I think people, I guess, will have an easier time recognising their own tasks, like for computer vision I can recognise my tasks in this way, and also to a certain extent generalise beyond their specific tasks, and understand it is like playing with lego blocks and combining them in different ways will lead to different results. If they will not see that in diagrams, I am not sure they would grasp this concept of compositionality as well through text. I don’t think they would do that.

Interviewer: So if you were describing some sort of NN system, how would you describe that in a diagram?

Participant: The second paper I told you has some examples of that. Usually what I prefer to do is not to be extremely detailed. For example there are some computer vision papers that I’ve seen people go all the way to each single tensor and all the dimensions of each single tensor, and that to me is a bit overwhelming. So what I prefer to do through the diagram is give the intuition rather than the specification. So usually the way I go and do that is usually by a flow diagram, where each single block, well it depends on the overall structure, but usually it is a layer or a cell in an RNN or something like that, and what I’m interested in trying to provide is how pieces connect to one another rather than a description of most of the pieces. Most of the time it is usually that the description of the pieces is not very important, it is just a hyperparameter configuration of the specific experiment, and you provide that in the experimental description or description of the protocol of the experiment, or even in the appendix. It is not a very core component. But what I try to do is show how this fits together.

Interviewer: And the diagram helps you to do that because?

Participant: Yeah, it gives lets say an overall description, while with text you’d have to give a description of each single component, and how they fit together, but maybe there are many connections and people can easily lose track of what connects to what and why, while a diagram gives something like a grounding or an anchor. Like “now I’m reading this description of a component but I forgot about the component before and how they were connected” and the diagram, just looking at it reminds you of these things, so it is an aid for understanding that, and the more detailed description you can get from the text, at least that’s how I use them. Other people may use them differently.

Interviewer: You described a level of granularity and abstraction that is of a certain type. How do you decide what to include?

Participant: Let me think about it. So usually [Pause] I would say one thing I try to do, in those diagrams that are specific for NNs, I would try to include an example. One specific instance, from the inputs and the outputs and what happens in between. Usually the way I decide to go in how to get there is the minimum amount of things that would make that specific example understandable, without being overdetailed and overwhelming, if that makes sense.

Interviewer: Yes, nice answer, thanks. And what tools do you use to make your diagrams, and why?

Participant: I think it depends. Usually, I don’t have a strong preference on tools, whatever allows me to do flow charts would work. Usually it is [Pause] I’m trying to remember. Lucidchart, that was one that I used, because you know the tool has a nice way to collaborate with other researchers. That’s actually the main reason that is useful. And the other could be just google docs, the presentation tool in google docs. Most of the time you will have to present those things too, and so you already have them and it also has collaboration tools attached to it, which is also great. Before I was using also omnigraffle, which has more tooling for the diagrams, but has the problem that it is more complicated for collaborating with people, because there is no easy way to share an online document, you have to send back and forth the documents, and that never scales.

Interviewer: Cool. So what sort of problems do you encounter when you are making diagrams about your research?

Participant: On one hand, balancing out the complexity of it is not easy. Particularly when you have a model that is big and complicated. Like if you look at the first paper I told you, that one has a huge diagram that is really complicated. So we tried to colour the diagram in a way that makes them independent and self-contained. But that one is really complicated, so. Balancing out how complicated it should be with how much detail it should have, and how much space do you have and how nice to look at it is and how understandable it is, that is one aspect for sure. And I guess the other difficulty is aligning things. Sometimes there are things. I feel like for CV it is sometimes a little bit easier, because in most cases those models have fixed dimensions, so they know exactly how many things should be there in the graph and in the diagram showing it. While for NLP models it is a bit more complicated, because there is usually at least one dimension that has a variable number of items, like the length of the text or something like that. And that system is even more complicated because you have also a variable number of terms, not just the variable number in each single term. So there is a second level of complication there. Making it so it can account for variable length things is tricky, when doing a diagram. Usually you have these dot dot dot things to make it so the user or the reader imagines a recurrent process or something. But it is a work around because you didn’t want to, or didn’t have the space or the patience to show the full process.

Interviewer: And how do you use diagrams when you are consuming research? Reading a paper, for example.

Participant: So I think that really depends. Again, different researchers have different to an extent styles of using a diagram. Some actually use them as the full description of their model. Others, more like me, use it as a way of grounding the whole conversation that is happening in the text. So depending on that [Pause] if I cannot find a description of the model anywhere else other than in the diagram, then I use the diagram as a reference. Usually I read through the text and then look at the diagram and use it as a confirmation that I am understanding things correctly. If I’m forgetting I don’t know “there was a symbol that was introduced at the beginning of the description and now later on the symbol pops up and I don’t know what it is, sometimes I look at the diagram and see oh yes there is this symbol which comes from this part of the model” or something like that. As a reminder or summary that I always use as a reminder really, of what I already read. A summary of that, yeah.

Interviewer: So when you are reading a paper for the first time, you start at the beginning and work your way through?

Participant: Yeah.

Interviewer: And if there is a diagram you will use the diagram?

Participant: So usually, sometimes I look at the diagram to get an overall view. And in particular when reading the model part. This I do when I reach the part that describes the model. Usually there are introduction and references before, so after that, then I look at the diagram. This is when I’m looking at the paper in detail, when I really want to understand the full picture. I look at the diagram and I have an overall understanding. Then I start to read, and when there is something that feels weird or I don’t understand it at all, then I look at the diagram and then I’m like “oh this is how it fits”. Then I keep reading, and I go back every now and then to the diagram to see how it fits together. But in other cases where there are papers that are really, in a specific subfield that I really know a lot about, sometimes the diagram is all I have to look at in order to understand the paper in its entirety. So it depends also on my degree of interest in the paper and degree of understanding of the subfield in which the paper is.

Interviewer: So if you were looking at a paper to decide if you want to read it, how would you decide whether you wanted to read it?

Participant: Usually I read the abstract first, because for instance if you look on arXiv you have the abstract and that’s the way you decide whether to click on PDF or not, right [Laugh]. If there was a diagram there, I would definitely use it as an additional tool to have, to make a decision. Probably not instead of the abstract, but together with this.

Interviewer: Yeah. I understand. If, for example, arXiv had the system diagram pulled out then that might be useful.

Participant: Yeah, definitely.

Interviewer: We’re probably not going to be able to do that as part of this work, spoiler alert! [Laugh]. So now you’re on arXiv, you’ve pressed the PDF button, are you now going to read the whole paper, you are committed at this point?

Participant: Well it depends [Laugh]. Usually I first skim through it, I look at the. Usually I skip the introduction. It depends on the subfield. If I am really confident on the subfield I skip the introduction and related work, as I most likely already know about them. I read a little bit about the model, look at the diagram. If this thing looks interesting, if sometimes I understand the diagram 100%, then maybe I don’t read the paper because it doesn’t look like something new to me. But if there is something weird and new in the diagram then I might start to look deeper in the methodology, and try to figure out what that thing that I didn’t know is, and maybe I can use it in my own future work, incase. So you know, I can say to a certain extent that I also use the diagram as a second step of filtering when I decide if I want to read a paper or not.

Interviewer: Do you look at things like F1 score?

Participant: I think it really depends, if I know a lot about the field, and I know there is let’s say a certain level of performance and the F1 score is really low with respect to what I know to be the current state of the art, then I am kind of suspicious. If it is too high, I am also kind of suspicious. But I’m not sure I use that as a kind of discriminator to me. I may be wrong about it, but evaluation is basically a way to confirm that whatever the methodology is is not completely flawed. But if the methodology is not completely flawed and doesn’t present anything new, it just improves 2% on the F1 score, but does not provide any new insight, either in training, the way it is trained, or in the architecture of the model, then I am not that interested in reading. I try not to read things that are only incremental, I try to read things that are also stimulating, interesting reads. Judging only on the values of the measures is not the best way to find those papers that are interesting.

Interviewer: I’m aware of time, we’re half way through already, so we’ll skip on to the example diagrams.

Participant: Yes, I have it.

Interviewer: What would be good is if you could tell me what you like and don’t like about it.

Participant: Alright. So this first one, Genre aware attention model, alright. [Pause]. For instance, for the first one, one thing I don’t like is the inputs. X1, xk, xn. I don’t really know, just from looking at the diagram, without a description coming from the rest of the paper, what k and n are. Because I don’t know if these are different sentences or what. It’s genre aware, it is NLP, so I guess the inputs are text or pieces of text, but I don’t know if they are words or sentences or anything else. So that doesn’t give me the full understanding of what happens there. So that I don’t like that much. Also that embedding layer that looks like a matrix is quite confusing because it is not really. What actually happens in that embedding layer is not really well reflected by that. The genre becomes a matrix and becomes a vector. There is an operation that multiplies by the matrix, so that is not really well depicted there. But the rest is pretty, I guess, straightforward and understandable. [Pause] Yeah, apart from what the inputs are, I think the rest of the diagram looks ok-ish.

Interviewer: So do you think inputs should be in the diagram? Should it be self-contained?

Participant: I would say. I would rather prefer an example of those inputs. Say if x1 was a word and xk was a word, I would understand that these inputs are all words. If the input was a sentence, even an example sentence taken from their dataset, that they are using for their experiment or something like that. That would be more clear to me than just xn, xm. The same thing is also true at the end. They have this sigmoid, but I don’t know what it is predicting, and what would be the ground truth here. So maybe if they added that too, to this graph, I could have understood the whole model and the whole task better. This way it is only the model that is described, but not the task. So maybe a bit more information about the task. It would have been easy also, to provide that sort of information.

Interviewer: You’re keen on having examples both in your own work and in the work of others. Do you reason about these systems by doing these operations in your head?

Participant: Personally yes. I use those examples to run through the models, to have some grounding. For instance at the moment in time where you have this green “score” box, and that output, what is that output. What does that output mean. If I know it is related to a sentence, then maybe that is a score of how this genre is similar to the sentence for instance. And that makes sense to me. But if I don’t know what x is, I don’t know what the score is, and I cannot reason down the line what all these arrows mean and what those operations mean. So for me, grounding is really important.

Interviewer: Great. So two funny little questions. Do you think the four and five circles are significant?

Participant: I see what you mean. [Pause]. I think no, I don’t think they are. I think the reason why they have them is because they want to differentiate between some blocks that are operations like “dense” and “score”, and some blocks that actually represent intermediate activation vectors. And so they decided that those dots are transmitting that information. And the fact that they have four and five, I think probably the reason why they did that was to show that the dimensions don’t have to match. But I don’t think it is that important, to be honest. They could have had another way to represent the fact that those are vectors rather than operations, like a different colour or I don’t know.

Interviewer: Speaking of colour, what are your thoughts on the use of colour in this diagram?

Participant: So I understand why they are using a different colour for the genre part, and the arrows and everything. But all the rest, I don’t think it is that important to have different colours. Also for instance I don’t understand why the different, I guess those are sentences, inputs, have different colours to be honest. Like pink, red and blue, like turquoise, whatever colour that was. I don’t think there is any use for that, and also score and softmax, why green and blue, there is no reason.

Interviewer: Cool. Does it annoy you, things like the colour being different for no particular reason, or is it fine?

Participant: Yeah sure. I think it is distracting to be honest. When I do these diagrams, I try to be mindful of the use of colour, to give it a meaning. Otherwise, why use it in the beginning? If it has no meaning, there is no use for that, to be honest. So it annoys me a little bit because it is making me think that those things are different while probably they are not. Why some operations are grey, why some operations are blue, some operations are green. It is confusing for me. If they were all the same colour it would be easier for me.

Interviewer: And overall, do you like or not like this one?

Participant: It is not that bad, I think it could be better with a better use of colour and the things I described before like inputs and outputs. But overall I think I have a decent understanding of what this model is doing, so I don’t dislike it. Not a superfan of it either.

Interviewer: Next one?

Participant: [B Pause] Alright, so this is a CV one, I guess. Ok so train a CNN. This is a bit small, the text. First person view, third person view. Ok. So about this, I think it is probably over-complicated for what it is trying to do. For what I believe I understand about what the model is doing. Because there is basically. My understanding is there are four input images, divided into two groups, and after several layers of encoding their output activations are concatenated and mapped in an embedding space with contrastive loss. Probably this could be in a much more simple way, without all these little numbers for the number of channels that is just a hyperparameters. Also the fact that this is six layers is also a hyperparameter, so it doesn’t really matter to me. I think it is a little bit too detailed, and the level of detail is not adding a lot. Also these little red arrows are not adding a lot. [Pause]. Yeah so this is the general feedback. I think I understand in general kind of what is going on here, particularly if I use the caption also. The caption is also helping. But yeah, I think I understand what is going on. It is probably too much, but still understandable.

Interviewer: What’s your feeling on the contrastive loss representation?

Participant: Because I know what contrastive loss is, I think to a certain extent it is ok-ish. I think probably the problem is that usually in contrastive loss you also have a negative sample, and here most likely they are only showing two positive samples, so it is not telling the full story of what contrastive loss is, but is just. To someone who does not know what contrastive loss is, that will mean nothing, basically. But if you know what contrastive loss is you kind of can understand what is going on, I guess.

Interviewer: And so overall do you like or not like this one?

Participant: Not particularly, I think it could have been better if it was simpler, like not showing all the number of layers and the dimensions of all the layers. It could also have been more generic, to an extent, because those are just hyperparameters and they don’t need to be in the graph, in my opinion.

Interviewer: Next one, C?

Participant: [Pause] Alright so this is more complicated, let’s try to look at it. [Pause]. Alright, so this one I have quite some difficulty in understanding it. I guess there are many terms that I don’t really know what they mean, and I guess that is part of the reason I’m finding it hard to understand it. And also the diagram is not really helping me a lot in doing that. So for instance, here, I don’t know exactly what guidenet is and what guidenet does. I can imagine that what this guidenet does is like colouring an image in a way that is useful, like classifying what are the cars or maybe the distance, most likely, how distant are the different elements in the stereo images, I guess. And I don’t really know what iteration zero, one and two are in the bottom right. And so I guess there is this guidenet producing an image DG and then that is fused with this cloud, I don’t know what that is, I guess it is a point cloud obtained by lidar I can imagine, but it is not really specified here. And why this model one, I don’t really know, and these single clouds, I don’t know, they are fused, but I don’t really know how, what they are. The image from guidenet is then combined with the clouds somehow, it is not really specified how. And then I cannot really understand iteration one and iteration two, I guess they are iterative refinements to also incorporate information from the cloud, but I cannot really say more than that from this graph.

Interviewer: So overall do you like it or not like it?

Participant: I would say no, because I don’t have a full understanding of what is going on, and the operations are not clear to me. So I would say no.

Interviewer: So D?

Participant: Alright so let me check this [Pause]. This may be from a paper that I know. This is probably done by a friend of mine [Laugh]. [Pause]. This was work from guys at [redacted] if I remember correctly [Pause]. Alright, so I think I have an overall understanding of what is going on. Again, that is most likely because I have a NL background, because otherwise I am not sure this graph is super super clear. I think I understand what the model does and what the goal of this model is, and why they are using colours. The only thing I don’t really understand is those little boxes, at the decoder level, the one with memory and schema. Those ones I am not really sure what they represent, but everything else I believe I understand quite nicely, I guess. Do you have specific questions for this one?

Interviewer: No, because you’ve already commented. The example is the interesting thing in this one for me.

Participant: Yeah, that example helps me a lot, I understand each single line. The fact it has the example and also it has the counterpart outputs for some specific items of the example and also for other items, those things that are coming out from each single green rectangle, those ones help me a lot in grounding. Like I know the fifth element maps to the fifth element in green, so because of that I know that it is trying to figure out what column, probably in a database, that word refers to, and the word it refers to is probably book title and years, so both of them are used there, so those are also columns. It helps me a lot in grounding it. If those were just x and y’s, I would not understand anything of what is going on.

Interviewer: So overall do you like or not like?

Participant: Overall I like it, yes.

Interviewer: Cool, E?

Participant: [Pause] So this one I can start to tell you I don’t like that some things are written horizontally, some things vertically, because I have to turn my head around to actually read them, which is kind of not super nice. [Pause]. And also the caption here is super non-informative too. I’ve done that too, to be honest. It also depends on the text, like maybe the text is super descriptive so you don’t want to have a caption that is exactly the same as the text in the paper, so maybe it could be the case there. Ok so I think I have an overall understanding of what is going on here. There are some things that are not clear to me, like the triangular thing in the procedure extraction module, it is not really clear to me what is going on there, to be honest. And also the use of dots, like lines close with dots, I guess it means there are tags associated with different parts of the video and transcript, because it starts and ends. But one block is used vertically, in the procedure extraction module, but in the procedure capturing module it looks like it is horizontally, which is kind of confusing to me to be honest. Personally what I would have done here, I would have made a more abstract graph here, of how the pieces fit together, and I would have detailed each individual element independently. But a similar thing can be said of my graph in [redacted] so I am probably guilty too of this same problem. Some pieces are not immediately understandable, like in the context aware fusion module those bubbles, those circles, I am not really sure what they are. I guess they are a bi-directional RNN but it is not explicitly said so, I cannot be sure of that just from the graph.

Interviewer: Do you think there are 10 or 20 of them, or is it just random again?

Participant: I think that this is as many as the inputs. It is bi-directional so it is probably 10 in each direction. But if you look at the transcript embedding one, there is self-attention layer n, for instance. They could’ve done the same thing at the top and just put bi-directional LSTM and just put one rectangle without background colour there, and achieved the same thing without all these circles which are confusing. If that is what they are doing, because I am not 100% sure what they are doing. Whilst with this self-attention layer n at least I know exactly what is going on. So I partially like the colouring sometimes. For some parts yes, for some parts no. For instance, for the context aware fusion module, I like the colouring because it is basically showing me that one part of the input comes from this green part which is from the video, I guess, and the dark green part comes from the transcript, because I know that both things are concatenated here. That specific use of colour I like. I’m not sure in the procedure extraction module what they are, I have no clue what it is. And in the last block the use of colour, green and dashed green, that actually probably is ok, is start and end so whatever is outside of those like sub-parts of the output which are tagged with something, which I believe is [Pause] predicted procedures, so they figure out that in that part of the input there was a procedure going on, so probably when there is no procedure those dotted dashed lines, those blocks are telling you something. That use of colour makes some sense. They are not using those inputs, they’re showing you that they are there but they are not using them, so it is kind of ok. [Pause] But I don’t know why some lines are black, some lines are blue. That I’m not really sure. Some use of colour is a little bit confusing.

Interviewer: Do you like the modules? I know you’ve said to pull them out into a more schematic one, but does the fencing of things, the dotted lines with the labels, does that work for you?

Participant: I think that works for me. One problem could be with that, that some parts are coloured in the same way, like context aware fusion module and transcript embedding are coloured with the same colour I guess, and I don’t know why. The others are different, and that makes sense, but those are the same colour. But in general the fencing things could be ok, but whatever is going on inside should be understandable on its own, as long as you have the inputs and outputs. And in this case that is not entirely true.

Interviewer: And overall do you like or not like this one?

Participant: I’m on the fence with this, because there are some things I like, some things I don’t. Overall, I’m slightly not liking it. The part on the left is pretty clear and I understand it I think, almost I guess. The part on the right is a little more confusing to me so I would say slightly no.

Interviewer: And F, the last of these examples?

Participant: This is actually more similar to the way I do my diagrams, like at this level of abstraction. [Pause]. Yeah so this one I like. I don’t know if it just because it is the way I do things, but I like it because it is entirely clear what the model does, to me, and it is entirely clear what the task is. There is a reading text, there is a summary, and there is a score at the end. To me this is perfectly clear. I think there are some little problems here. I don’t understand why some blocks are round, some blocks are of different sizes, some blocks are dashed, and I think they could all look the same without [Pause]. It would probably be better to make them all look the same. Unlike some of the other graphs, where some blocks were activations and some blocks were operations, here they are all operations, so I don’t really see why. The only thing I don’t really understand is merge, because it is not really explained, at least from the graph, what that merge is. It could be the case that in the text they say merge could be several different things and they tried several of them, so that could make sense to have a generic merge operation, which is actually similar to what I did with [redacted]. I do like this one.

Interviewer: I’m aware we are at time, how are you for time?

Participant: I can have a few more minutes if I switch rooms.

Interviewer: That would be really amazing.

Participant: Ok just give me a minute while I switch rooms [interruption]

Interviewer: So you like this last one. So we can probably move on to the next part, about tasks. We started out doing ranking, but it turned out to be too much, so if you could just pick the top 2 or 3 you think are important, and any that you really don’t do and are not a use case for you.

Participant: I see, I see. The first one I’d say yes, I use it for. I don’t know if it is the top, let’s read the other ones. The second also true, relations between components. External dependencies, not really, it is not something I usually look for. Corpora and data types, it depends. Probably not corpora, but data types maybe yes, depending on the different graphs, it could be yes. Representational choices isn’t something I usually do. Purpose of the system I would want to use it for, but it depends, in some cases I can do that if inputs and outputs are specified, in other cases I cannot do that, but it is something I would want in my graphs, and my diagrams to show. Specific architectural features, definitely. Opportunities to alter [Pause] probably not, I wouldn’t do it from the diagram directly. It is probably more something about the model itself, once I understand about the model then I can understand if I want to alter it, but not only using the graph. [Pause]. Identifying author intention, no. Usually that’s true, obviously what the author puts there is what they believe is important. But at the same time I don’t always agree with what the author thinks is important. So this point isn’t super important to me. Comparing to other systems, not really, because there is no single way. Even the same system can be represented several different ways at different levels of abstraction, so sometimes it is not straightforward or easy to compare to other systems, just from the graph. Sometimes authors, to compare to other systems, redesign the old system with their own graphical notation. In that case you can do that, but otherwise you cannot. [Pause]. Yeah initial check for sure. Index definitely yes, memory aid also. Aid for writing a summary, not really. How the system works, at a coarse grained level, most likely yes. Ok I have to say the top three and bottom three. Index to navigate the paper is in the top, memory aid also in the top, and to be honest understand how the system works is also in the top but is not that different from identifying system novelty and contribution, because if I understand how it works I understand the novelty. But I would say understand novelty at a general level rather than the specifics probably yes. And the bottom one I would say external dependencies, what the author thinks is important, and the summary, those are my bottom ones.

Interviewer: Is there anything not on that list that you think is an important use of diagrams for you?

Participant: No, I think that captures most of the things. [Pause] I believe that captures most of the things, certainly the things I use it for for sure.

Interviewer: In terms of extracting an example, it sounded like you were using the diagram to do that. Is that a distinct thing in your cognitive processing, that example manipulation?

Participant: For me no. These things are not separate. Without the example helping me grounding, even not a super specific one, just what are the inputs what are the outputs, without that I wouldn’t know what some of the pieces are. How to interpret what the inputs and outputs to some of the pieces would look like or what they would be like. If I’m not able to do that, I am not able to really understand the system. Usually in the textual description, you are forced for good to use some symbolic representation, to use some variables to define the intermediate states, and the inputs and the outputs, but I think the diagram is a useful piece for grounding that into an example, so that you understand sometimes if you sometimes have attention maps you have these graphs that are attention maps that are showing you the model and focuses on the input part of another one. If you have that, that helps you understand things quite well. So I wouldn’t call them entirely separate. They are two separate things but they work together in the entire diagram understanding.

Interviewer: Yep, got it, I think I understand. The last part is some broader questions. Do you use diagrams to reason as part of your creative process when doing research? And if so how?

Participant: Yes 100%. So usually the first step in a new research idea for me is drawing a diagram on the whiteboard. Then the second step, when I have overall idea, the diagram helps me guiding my design process, and when I have the general idea in my mind through the diagram, then I start to actually use math to say “ok this piece of the diagram is supposed to do this, let’s write the math, so what actually happens and get more details”. Diagrams for me are the first step, and the easier way to convey things, also to people who are non-researchers, like product managers, these kinds of people. It’s the easiest way to convey what I have in mind.

Interviewer: You use mathematics rather than code as the next level down?

Participant: Yes, in particular if I work on a whiteboard and I want a generic understanding of things. Then when I want to implement things obviously I use code. I think I almost never, it is really rare for me to write code on the whiteboard, or pseudocode, because the pseudocode can to a certain extent be represented by the graph itself, and the details can be observed by the maths, and those are to some extent higher, more abstract languages than the code. I only get to the code when I actually have to write it.

Interviewer: Cool. And when you’re writing the code, do you interact at all with the diagram?

Participant: It depends. Usually I use it as a memory aid, but not much more than that. I probably interact much more with the math, because if I broke down the math somehow then I try to implement that math in the model, so that is more direct. It is more detailed than the diagram, so when I go to implement I go directly to the math and not to the diagram. The diagram could help me in organising the structure of the code, to a certain extent, but not in the implementation of a specific function. The math part helps me more with that.

Interviewer: And so you’re doing it on a whiteboard, are you doing it with other people?

Participant: Usually yes, it depends on the different projects. Sometimes I have an idea and just sketch it, not bouncing it off anyone. Sometimes it is a brainstorm session, and we have a new task that needs to be done or there is a basic paper idea coming from someone, and we write it collaboratively on a whiteboard. Yeah. So usually we try to use different colours, so we can remember the different things that others have said, to keep track, like on google docs when you edit, something like that [Laughs].

Interviewer: Do you use diagrams in any other way during your research more broadly?

Participant: [Pause] Because I also have this angle of system design, particularly with [redacted], I have to structure the code and the work around [redacted], I use diagrams to keep track of the moving pieces of the systems. I did most of the work on it, but there are also other people collaborating, so to keep track of who is working on which part I usually also use diagrams for that. And also something like Gannt diagram to track tasks over time, to align work with different people collaborating on the same project.

Interviewer: You don’t have product roadmaps?

Participant: Again it depends, because most of the efforts that we do are relatively small teams, so it is not really product people. Projects that end up being developed in products or added to products, those have product managers who actually that’s their job to keep it organised. For more researchy projects, there are only two or three people at most. We are usually a bit less structured than that. We use some diagrams just to keep track of our work, but not really heavy tooling and heavy organisation I would say. It is more informal, kind of like a university to be honest. There are some research groups which are more organised, but I think you see what I mean.

Interviewer: Yep. You use block diagrams a lot, with words in boxes with lines. Do you use them exclusively or are you inspired by other’s work and might be inspired to replicate bits of what they’ve done diagrammatically?

Participant: I’m not religious about it I guess [Laugh]. If I see something that works, I can definitely end up using it. There is no [Pause]. It is not only about blocks, I can use anything. Usually I tend to use the simplest thing possible. So for instance I try not to use tensors that are shaped like cubes and things, because it is just adding complexity in most cases and is not useful. In some cases it may be useful, in that CV paper we had if they just had one diagram then in that case it would be super useful to have the dimensions. But it is probably the exception, not the rule. And regarding what other people do, I’d say yeah I can definitely be inspired by what other people do. But to a certain extent, I haven’t seen many [Pause] new things coming out from people recently, having better ways to represent things. So I would say one thing that I try to do and take inspiration from, and they’re not even researchers most of the time, is using aligning things, using fonts, consistency in colour and shape and size, but they’re not really. The inspiration doesn’t come from other people doing diagrams in research, more from generic design principles. I try to learn more and more, even if it is not my job. So I would say grab an influence but not usually from other people’s graphs but more from generic principles of design.

Interviewer: That leads us on quite nicely to the last question, which is are there any tools that you think might help you in creating or consuming diagrams? Tools can be cognitive, physical or virtual tools.

Participant: I see. Regarding cognitive tools, again there is nothing that I know of at least. There could be things out, like frameworks or something like that. Of how to do things that I might not be aware of. Something like that, some guidelines, or a set of guidance, something like that, could be super useful for researchers because most people don’t really know what they are doing and don’t know even basic things about use of shape and colour and fonts. Having some guidelines around that, and maybe also some oh yeah just one step back. In terms of inspiration there is one person I draw inspiration from sometimes, because I think he does really good diagrams, and that is [redacted]. He’s great. Whenever he puts something in his graphs, that something has a meaning, and it explains what that meaning is and why he is doing that. So I think his graphs are amazing. Back to the framework. Something like he described for depicting RNNs, those kind of guidelines could be super useful. Every time you have an arrow, that arrow means something. Every time you have a box, that box means something. Every time you use a colour, that colour means something. Something like a framework like that, that people could decide to follow or not, and at least draw some inspiration from, could be useful. But I don’t know anything that has been done about it. Regarding physical tools, I’m not sure. I’d say if there was a set of guidelines like those, maybe a physical tool that makes it easy or by default follows those guidelines, that would be great to have. For instance if you decide that just to make a super simple example, if you decide that all operations have round shape, and all have the same size, maybe a tool that allows you, maybe if you have a description that is bigger and overflows and you resize that box to fit the new description, maybe it resizes all those boxes that are semantically similar or semantically the same. But in order to be able to have a theory or a framework behind that, that tells you what to do. A physical tool that enables me to do that would be useful, and also maybe I guess something like that would be useful. And I guess maybe you take a picture of your graphs that you did on a whiteboard and it translates it into a graph that you can already display on your computer, that would be also super useful as it would make me do less work. Less redoing the same work on the computer that I have already done on a whiteboard. That would be super useful.

Interviewer: I think you’re not the first person to mention liking that blog. I haven’t seen many people copying that style in conference proceedings. Why do you think that is?

Participant: Maybe one possible explanation could be that, in particular for the RNNs one which is the most referenced, I would say, it is because people look to a higher level of abstraction. Now, the internals of an RNN cell are known, so people know what they are, so people are stacking several on each other, these kinds of things, and so they don’t need. Once the description of the inner workings of that thing is known you don’t need to have the description of that thing any more at all. So you can move to a higher level, to a certain extent. That could be a reason I guess. It was also not presented in a paper itself, as a thing you can reference. If someone proposed that thing, or something similar to that, as a paper and as something people might want to use, and goes to a conference on machine learning and does a presentation about it, then people are aware of it. Maybe that could stick more. And also probably because there was no tool around it. If I want to replicate what he did, I cannot replicate it 100% because he didn’t say what tool he was using for doing that. Probably it is a friction of not being easy to use it, and to be able to replicate it, I guess.

[The interview finished with some informal discussion about the future direction of this work]