Personal Dynamic Memories are Necessary to Deal with Meaning and Understanding in Human-Centric AI

Luc Steels[1]

Abstract. Human-centric AI requires not only a fundamental shift in the way AI systems are conceived and designed but also a reorientation in basic research in order to figure out how AI can come to grips with meaning and understanding. Meanings are made up of distinctions to categorize and conceptualize an experience at different levels, from directly observable factual meanings to expressional, social, conventional and intrinsic meanings. Meanings get organised into larger-scale narratives that conceptualize experiences from a particular perspective. Understanding is the process of constructing and then integrating these narratives into a Personal Dynamic Memory that stores narratives from past experiences. This memory plays a crucial role to construct more narratives and thus works intimately together with inferences, mental simulations, and the analysis of experiences in terms of syntactic and semantic structures.

This paper outlines this approach to meaning and understanding by clarifying what it entails, outlining technical challenges that must be overcome, and providing links to earlier relevant AI work as well as new technical advances that could make Personal Dynamic Memories a reality in the near future. [2]

1 What is human-centric AI?

"Human-centric AI focuses on collaborating with humans, enhancing human capabilities, and empowering humans to better achieve their goals." [17]. Human-centric AI has become a focal point of current research, particularly in Europe, where it is now the stated objective of the EU strategy recently (February 2020) issued by the European Commission. This strategy calls for AI that shows human agency and oversight, technical robustness and safety, privacy and data governance, transparency, care for diversity, non-discrimination and fairness, focus on societal and environmental well-being, and accountability [36].

Achieving human-centric AI requires a number of changes in focus compared to current AI:
(i) Human-centric AI systems should be made aware of the *goals and intentions* of their users and base their own goals and dialog on *meanings* rather than on statistical patterns of past behavior only, even if statistical patterns can play a very important role, for example for drastically reducing search or carrying out approximate inference. Human goals and values should always take precedence. Respect for human autonomy should be built into the system by design, leading to qualities such as fairness and respect.
(ii) Human-centric AI requires that a system is able to explain its reasoning and learning strategies so that the *decisions are understandable by humans*. Only by emphasizing human understandability will human-centric AI achieve proper explainability and transparency.
(iii) Human-centric AI should not only learn by observation or theorizing about reality but also by *taking advice* from humans, as suggested in John McCarthy's original 1958 proposal of the Advice Taker [13].
(iv) Human-centric AI should be able to use *natural communication*, i.e. communication primarily based on human language, not only by mimicking language syntax but, more importantly, using the rich semantics of natural languages, augmented with multi-modal communication channels. This is needed to support explainability, and accountability.
(v) Human-centric AI should have the capacity of *self-reflection* which can be achieved by a meta-level architecture that is able to track decision-making and intervene by catching failures and repairing them. By extension, the architecture should support the construction of a theory of mind of other agents, i.e. how they see the world, what their motivations and intentions are, and what knowledge they are using or lacking. Only through this capacity can AI achieve intelligent cooperation and adequate explicability, and learn efficiently through cultural transmission.
(vi) Finally, human-centric AI should reflect the *ethical and moral standards* that are also expected from humans or organisations in our society, particularly for supporting tasks that are close to human activity and interest.

Today the dominating perspective on AI is not human-centric. It focuses primarily on achieving high predictive performance on predefined benchmarks, trying to exceed human performance so that humans can be replaced in the task being considered. This approach is *machine-centric* rather than human-centric. It emphasizes numerical (subsymbolic) techniques (from neural network research, pattern recognition, information retrieval, and data science), often ignoring valuable contributions from symbolic AI that are needed to achieve explicability and robustness.

Admittedly the machine-oriented focus has recently lead to a jump in performance on chosen benchmarks, particularly in the domain of pattern recognition and computer vision, but unfortunately also to a kind of AI that is opaque, cannot explain or defend its decisions, is unable to take human advice, is not robust against adversarial attacks, has no understanding of the motivations of its users, and requires vast amounts of data and computing power. Although for a large, growing class of applications these shortcomings are not an issue, for AI ap-

[1] Catalan Institute for Advanced Studies ICREA - Institute for Evolutionary Biology (UPF-CSIC) Barcelona Spain, email: steels@arti.vub.ac.be

plications that touch on human lives and are socially consequential, these disadvantages are highly problematic.

Different approaches to human-centric AI have been proposed recently. They are all valuable. Some researchers have advocated guidelines and design methodologies to make AI more trust-worthy and responsible by emphasizing safety, privacy, data governance, transparency, diversity, fairness, and accountability [30], [7]. Others have emphasized that we need more human-centric interfaces for AI systems, including better explanation facilities and ways for humans to provide guidance during machine learning or decision-making[38].

Here I focus on the idea that *human-centric AI requires above all another kind of AI*, namely AI which has meaning and understanding at its core. The present paper is a position paper, trying to clarifying this point of view and reflecting on the key issues and possible technical solutions. But first, what do we mean by meaning and understanding?

2 Meaning and understanding

The notion of meaning is related to how we try to understand how humans make sense of an experience. An experience can be a behavior or the observation of a behavior, an image or a sequence of images, sounds, soundscapes, smells and tastes, spoken or written text, and more generally cultural artefacts like scenes in a theatre play. In the real world, there is a flow of experiences that we need to interpret and cope with quickly. For example, if we are driving a car there is a quick succession of situations that we have to gauge correctly in order to act appropriately, even in unusual situations: Why is the car behind mine honking its horn? Is the woman with a baby stroller going to cross the street or has she seen me coming? Why is everybody slowing down? What does this red light on the dashboard mean?

Meanings are built from categorisations of reality, for example, colors, actions types, temporal and spatial relations, etc. Categorisations are distinctions that are relevant for the interaction between humans (or agents more generally) and their environment, including other agents [25]. For example, the distinction between red and green is relevant in traffic lights because it tells you whether it is safe to start driving or cross the road. The distinction between angry and sad is relevant for knowing how to behave with respect to another person. The distinction between left and right is relevant for giving or following instructions how to reach a location or how to find an object in a scene.

Categories are the building blocks for constructing different levels of meaning for an experience, The following levels are often discussed in the appreciation of art works [18] but are actually useful for interpreting any kind of experience [27]:

- The base level of an experience details the external *formal properties* directly derivable from the perceived appearance of the experience, for example, the lines, shapes, color differences in hue, value (brightness) and saturation, textures, shading, spatial positions of elements, etc. in the case of images.
- The first level of meaning is that of *factual meaning*. It identifies and categorises events, actors, entities and roles they play in events, as well as the temporal, spatial and causal relations between them. In the case of images they require a suite of sophisticated processing steps, starting from object segmentation, object location, object recognition, 3D reconstruction, tracking over time, etc.
- When there are actors involved, a second level, that of *expressional meaning* becomes relevant. It identifies the intentions, goals, interests, and motivations of the actors and their psychological states or the manner in which they carry out actions.
- The next level is that of *social meaning*. It is about the social relations between the actors and how the activities are integrated into the local community or the society as a whole.
- The fourth level is that of *conventional meaning*, based on figuring out what is depicted or spoken about and the historical or cultural context, which has to be learned from conversations or cultural artefacts, like books or films.
- The fifth level is known as the *intrinsic meaning* or *content* of an experience. It is about the ultimate motive of certain images or texts, or why somebody is carrying out a certain behavior. It explains why this particular experience may have occurred.

We define a *narrative* as a coherent reconstruction of the different levels of meaning of an experience or a set of experiences based on one or more perspectives. It contains categorised entities at each of these levels, links between the levels, and possibly additional cross-level categorisations. The perspective, which is often the perspective of the agent itself, is unavoidable because categories are most of the time observer-dependent. For example, an object which is to my left is for a person opposite of me to the right. I may categorise a gesture as aggressive whereas the person making the gesture may have performed it to defend herself. I may not know a particular historical figure and believe it is just the representation of an old man, whereas you may recognize the figure and be repulsed by the atrocities that were conducted under his command. Transforming a narrative from one perspective into a narrative for the same experience from another perspective is a critical component in handling meaning. Even to communicate properly in language we often have to look at the viewpoint of the interlocutor and categorise spatial and other relations accordingly.

Understanding is a process with three functions: (i) Reconstruct the different levels of meaning by casting them into coherent narratives that explain the events underlying the experience, (ii) predict how the experience will unfold in the future and reconstruct what has happened in the past and (iii) integrate these narratives into a *Personal Dynamic Memory*. A Personal Dynamic memory is an active store of past experiences which may include partly some of the original data but mostly the webs of meanings and the narratives that have been constructed during the interpretation of earlier experiences. A Personal Dynamic Memory is crucial for supporting the construction of narratives of new experiences but it is today missing from existing AI systems.

Here is a simple example to illustrate these ideas. Consider the image in Figure 1 (left). This is from a poster that used to be employed in French and Belgian schools to teach children about daily life and to learn how to talk about it. We instantly recognize that this is a scene from a restaurant, using cues like the dress and activities of the waiter and waitress or the fact that people are sitting at different tables in the room. Current image recognition algorithms would be able to segment and identify some of the people and objects in the scene and in some cases label them with a fair degree of accuracy, see Fig. 1 (right).

However a normal observer would see a lot more than that. For example, when asked whether a person is missing at the table on the right, the answer would be straightforward: Yes, because there is an empty chair, a plate and cutlery on the table section in front of the chair, and a napkin hanging over the chair. So there must have been a third person sitting there, probably the mother of the child. Moreover nobody has a lot of difficulty to imagine where she went.

There is a door marked ‘lavabo’ (meaning ‘toilet’ in French) and it is quite plausible that she went to the toilet while waiting for the meal to arrive. Any human viewer would furthermore guess without hesitation why the child is showing his plate to the waitress arriving with the food and why the person to the left of the child (from our perspective) is probably the father looking contently at the child. We could go on further completing the narrative, for example, ask why the cat at the feet of the waitress looks eagerly at the food, observe that the food contains chicken with potatoes, notice that it looks quite windy outside, that the vegetation suggests some place in the south of France, and so on.

Fig. 1. Left. Didactic image of a scene in a restaurant. Right. Image segmentation identifying regions that contain people (based on Google’s Cloud Vision API).

Clearly these interpretations rely heavily on inferences reflecting knowledge about restaurants, families, needs and desires, roles played by people in restaurants (waiter, waitress, bar tender, cashier, customer). These inferences are not only necessary to properly interpret the visual image in Fig. 1 but also to answer questions such as ’Who is the waitress?’, ’Why is she approaching the table?’, ’Where is the missing person at the table?, ’Who will get food first?’, etc., We can also make predictions and reconstructions, for example, that the waitress will reach the table, put the food on the table, cut the chicken into pieces, and put them on the different plates, or that the mother of the child will come back from the toilet, sit down again at the table, and start eating herself.

Each of us has a vast Personal Dynamic Memory that stores narratives based on prior experiences: from visiting restaurants, seeing images in pictures or movies, reading about them, etc. Our daily life is filled from morning to evening with activities to feed and reorganise our Personal Dynamic Memories and the richer they become the more we are able to make sense of new experiences. What is truly amazing is that by the time we reach the adult stage these memories must already contain a massive number of facts, which are nevertheless searched at an incredibly fast rate with relevant parts of memory becoming primed and ready for use for handling novel experiences.

Understanding uses information both from syntactic and semantic parsing of the experience and from inferences based on a Personal Dynamic Memory, in order to fill in unexpressed or un-observable information, e.g. via logical reasoning and mental simulation. Moreover the understanding process changes the contents of Personal Dynamic Memory, not only because the new experience, its interpretation, and links to other experiences are stored, but also because earlier experiences are revisited and their storage may be affected by newer experiences. Memory needed for understanding is therefore highly dynamic, unlike computer memory that remains unchanged once something has been stored.

This leads to the proposal for a general architecture for AI systems that handle understanding depicted in Fig. 2. It shows the flow from experience to syntactic and semantic structures, and from there towards the construction of narratives, integrated into a Personal Dynamic Memory. The flow of information is not only bottom-up but also top-down, shown with the green arrows. The narrative under construction is partially guiding semantic analysis and cutting down combinatorial search in syntactic analysis, whereas the narratives already contained in the Personal Dynamic Memory are guiding the construction of narratives of new experiences.

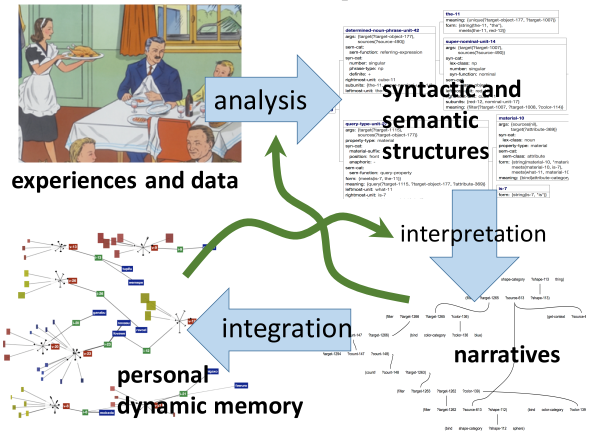

Fig. 2. Components for tackling understanding in AI systems. Besides the introduction of narratives, the main critical component is a Personal Dynamic Memory which helps to build narratives to interpret a new experience. The green arrows indicate that there is strong downward information flow from the Personal Dynamic Memory to the interpretation process and from narratives to the analysis process.

3 Current AI does not handle meaning properly

Before putting some more technical flesh on this architectural skeleton, I want to emphasize that current techniques and AI design methodologies are not handling meaning and understanding. Current techniques fall into two classes: numerical (or subsymbolic) techniques and symbolic techniques with shades in between.

Simplifying, numerical (or subsymbolic) AI techniques translate problems into a numerical form (real numbers and vectors) and perform numerical operations over them. The numerical representations are constructed using information-theoretic considerations, specifically, their ability to help predict or complete patterns. Most neural networks fall into this class, but also other techniques like Latent Distributional Semantics, which associates a vector representation known as an *embedding* with words, images, or actions. The embeddings capture the syntactic and semantic contexts in which an element appears and can be used to compute similarities, predict the next word or image, relate an image to a label, answer textual queries, or perform many other useful subfunctions for building intelligent applications. Embeddings are computed either by statistical methods or by using deep learning algorithms.

Importantly, and as pointed out clearly by Claude Shannon [24] who can be considered the father of numerical AI, information-theoretic representations do not try to capture meaning. For example, a word embedding captures the kinds of contexts in which a word may occur but this is only an indirect substitute for the real meaning of the word. Ignoring meaning makes it feasible to use these numerical techniques in circumstances where there is no representation of meaning available for learning or training - which is in fact almost always the case. But it leaves out a crucial aspect of (human) intelligence.

Thus, ‘Neural image labeling’ associates rather directly labels with images (sometimes even using only pixel-based image representations), without attempting to discern individual objects, actors, or events, and without trying to figure out the situation underlying the image, the nature of the action, the motivations of the actors depicted in an action, the historical setting, the reason why the image is made, and many other aspects which human viewers spontaneously come up with. 'Neural translation' does not try to perform a syntactic analysis using grammars and parsers nor semantic analysis using interpreters building conceptual representations of what is being said. Rather, they associate n-grams in the source language with n-grams in the target language based on word vectors that capture statistical co-occurrences in dual (source/target) corpora.

Circumventing meaning has made the current wave of deep learning based AI applications possible but it is also responsible for the brittleness of image labeling, the nonsensical nature of translations, failures in answering questions that fall slightly outside of the statistical patterns in the corpora used to train them, the success of adversarial attacks for interpreting images or texts that do not confuse humans but throw off AI systems, the non-transparency of decision making, and many other features that human-centric AI considers undesirable.

Intuitively a kind of *hybrid or integrated AI* that combines the virtues of numerical with those of symbolic AI is a possible way out and has indeed been proposed by several researchers. Symbolic AI maps problems into symbols and symbolic structures and performs transformations over these symbolic structures, for example guided by rules of sound logical inference. This approach flourished in the 1970s and 1980s leading to expert systems built for interactively supporting experts, large-scale ontologies and domain models as now used in the semantic web or in encyclopedic knowledge-graphs, computer-assisted theorem proving, constraint solvers for scheduling or design, precision language processing, and much more.

The symbolic approach has tried, at least in principle, to get closer to handling meaning. It has used terms like semantic information processing [14], or story understanding[22], talked about AI able to take advice, rather than be programmed explicitly or trained with large data sets[13], and built sophisticated explanation facilities for expert systems using deep human-comprehensible models of the domain and an explicit representation of the problem solving methods being used[16].

Nevertheless, the symbolic approach has its own limitations with respect to handling meaning and understanding. A key criticism, reflected in Searle's Chinese Room argument and known as the *symbol grounding problem*, is that symbolic AI operates in a world of symbols with no systematic connection to the real world. To solve this problem requires an integration of a symbolic and a numerical approach, because the latter starts from the (real) numbers delivered by sensors and actuators that are directly connected to the world, so that the categories that constitute the meaning of symbols indeed become properly grounded. However, it is important that the grounding of symbols is based on what is meaningful, i.e. relevant, to the agent, which is different from grounding based on success in prediction tasks. When agents cooperate on tasks in a shared environment, particularly if they have to communicate about tasks, they implicitly have to coordinate the way the categorize reality and how these categorizations are expressed.[28]

In addition, the transformations of symbolic structures are formal operations, similar to a set of axioms and rules of logical inference as in mathematics. But the problem is that it is very hard, if not impossible, to define axioms exhaustively for real world open-ended domains due to the unavoidable exceptions, lack of knowledge, and the problem of making clear-cut definitions. These problems have been discussed widely under the title of the *frame problem*. Also here an integration of numerical and symbolic techniques is a way to go forward so that the flexibility of pattern recognition and action selection based on neurally inspired models, which gives only approximate answers, can be married to the precision and compositionality of symbolic reasoning.

4 Relevant work

The ideas proposed here are certainly not new. For a long time it has been commonly accepted in cognitive science that the construction of narratives is an essential ingredient of cognitive intelligence because it allows us to make sense of reality [6] [35]. Also in AI there has been significant prior work, although mostly in the context of story generation and story understanding, which are the textual manifestations of internally constructed narratives [12]. We find symbolic approaches from the late nineteen-seventies onwards, such as in the work of Schank and colleagues [23], Winston's proposals for Computational Narrative Intelligence [37], or more recently the work of Gervas and his group on narrative generation [9]. There is also increasing work at the moment using numerical approaches towards narrative intelligence [20], particularly within the context of building question-answering and dialog systems.

In the psychological literature there has also been extensive work on personal memory, often based taking Tulving as a starting point [33]. He introduced the distinction between procedural (knowledge of skills) and declarative memory, usually divided into semantic memory, which contains general factual knowledge, and episodic memory, which refers to specific autobiographical experiences stored in the form of contextualized past perceptions, actions and temporal and causal structures. Schank and colleagues have made proposals in the late 1980s on how such dynamic memories could be built[22]. This has lead in the nineteen nineties to significant work on case-based reasoning [2] and memory-based reasoning [26]. Much of this has been overshadowed by the current peak of interest in deep learning, but it remains highly valuable for the aims discussed in this paper.

Meanwhile various important technological advances have been made in other areas that make a renewed effort towards the experimentation with Personal Dynamic Memories and narratives a realistic prospect. Among these advances I just want to highlight the following:

- *Very large knowledge bases.* One of the critical bottlenecks for effective Personal Dynamical Memories is the sheer size of the knowledge that has to be represented and processed. If we express this in terms of facts, then we must expect to handle at least tens of millions, if not billions. This was totally impossible two decades ago but very significant progress, pushed by the development of the semantic web, has changed the situation. It is now possible to represent fact-bases up to 100 billion triples using standard knowledge representations (RDF statements and OWL) and perform inferences over them fast enough to be used in interactive applications[34]. So the issue of computational complexity for Personal Dynamic Memories can be considered to be solved.
- *Robotic embodiment* Another critical bottleneck is that Personal Dynamic Memories have to be grounded in sensori-motor experiences. A few decades ago the state of the art in computer vision and robotics was simply not advanced enough to tackle this issue in any realistic way. But also here there have been tremendous advances, both in the availability of lower cost robotic hardware including cameras and signal processing chips and in software for

perception and action control, primarily using techniques from deep learning. These developments in themselves do not solve the issue of symbol grounding but they have made it possible to start addressing it seriously. One example of recent work uses language games between embodied autonomous robots that generate not only their own communication system but also an ontology containing the relevant distinctions in a specific domain [31]. These experiments have shown how perceptually grounded categories (for example for color or size) or spatial and temporal relations grounded in event recognition can emerge in populations of agents pushed by the task of communication. Another example is the Open-Ease framework `http://www.open-ease.org/` that supports the recording and storage of inhomogeneous interpretation data from robots and human manipulation episodes so that they can be used to build semantically oriented tools interpreting, analyzing, visualizing, and learning from these experience data.[4]

- *Mental simulation* Another bottleneck for building realistic Personal Dynamic Memories has been the role of mental simulations of actions and situations. This is considered an essential function of memory by many psychologists, particularly for predicting how a perceived situation will continue to evolve in the future[3]. This hypothesis has also inspired AI researchers[5] but implementations could only explore simple isolated examples until very recently. However, significant advances in virtual reality technology have now pushed the state of the art in computer graphics to allow a very high degree of realistic simulation even for complex world situations, thanks also to dedicated hardware (game engines who have now reached performances of 12 terraflops) and highly optimized software. This technology is already being used for cognitive robotics experiments in order to plan future behavior through mental simulation, complementary to classical planning based on symbol manipulation, and to understand human language instructions or descriptions.[19] So also for this aspect, there are promising developments that make Personal Dynamic Memories much more feasible.
- Finally there have been significant advances recently in Computational Construction Grammar. Most linguistic formalisms, such as Chomskyan generative grammar, remain close to the morphosyntactic structure of a language. *Construction Grammar* in contrast focuses on capturing the systematic ways in which grammar expresses meaning [11]. It is therefore a more appropriate basis for natural language processing for an AI approach that seeks to handle meaning and understanding, particularly because Construction Grammarians have worked closely with cognitive semantics [32], an approach to semantics that seeks to understand the conceptual patterns with which humans organise their experiences in order to make it expressable in their language. A decade ago usable implementations of construction grammar and cognitive semantics were in their infancy but this has changed completely. A first big effort, spearheaded by ICSI in Berkeley, developed an Embodied Construction Grammar[5], which not only formalized and operationalized construction grammars but also subscribed to the 'mental simulation' approach to meaning mentioned in the previous paragraphs. Another big effort, at the University of Brussels VUB AI Lab and the Sony Computer Science Laboratory in Paris, developed Fluid Construction Grammar[29], which has now a very solid implementation and a growing user community.(see `www.fcg.org`) Given that language communication plays a major role in the way that human Personal Dynamic Memories get formed, this line of research provides another hopeful contribution towards achieving meaningful AI.

5 The organisation of memory

In my opinion, the most critical bottleneck at the moment is: How should a Personal Dynamic Memory be organised at the micro-level and what kind of basic computations (including inferencing and learning) should be supported. Obviously a linear list of facts, possibly represented in RDF, will not do, we need higher level structuring devices, partly for managing inferential and combinatorial complexity, partly for dealing with the frame problem, and partly for achieving fast access to the most relevant prior experiences that will help to make sense of a new experience. What will also not work is to blindly store the vast amount of information generated by an experience, the complete sensori-motor data streams, the data from the mental simulations that are triggered, the language descriptions and their semantic interpretations, or all the facts relevant for an experience. If eveything is stored this is not only costly from an energetic point of view but will certainly get in the way of fast retrieval and inference.

The cognitive science and AI literature already contains various proposals for the organisation of memory. Many of them start from Bartlett's original idea of a *schema* also called a *frame*. It was formulated in the 1930s and revived again in the 1970s by psychologists such as David Rumelhart [21], linguists such as Charles Fillmore [8], and sociologists, such as Erwin Goffman [10].

A schema is a way of framing a particular situation in terms of a set of entities, roles for these entities, constraints on the kind of entities that can fill these roles, and relations between the entities based on their roles. Each schema has various associated cues to recognize quickly whether it applies to the current situation. Once it is triggered, a schema casts a web over the sensori-motor inputs and facts associated with the situation and it makes us see or infer certain aspects of the experience more clearly at the expense of others. Schemas impose a bias and perspective on a situation and often also an emotional reaction. They come with a lot of defaults. These are facts which can be expected to be the case if a particular schema matches well with an experience, but are not explicitly mentioned or observable. Sometimes these defaults even override perception or fly in the face of obvious facts.

The notion of a schema was introduced into AI by Minsky[15] who used the term frame. It lead to a variety of frame-based knowledge representation systems in the 1970s, which were used extensively to model the perception of complex scenes, story telling and story understanding, and expert reasoning. Frame-based representation systems feature datastructures for representing frames, basic inference operations over frames, and languages and interfaces to define frames and maintain large collections of frames. Frame-based knowledge representation systems also support various kinds of relations between frames, in particular subtype relations so that there could be the inheritance of information from one frame handling a broad set of experiences to another frame concerned with a more specific situation. Another example are priming relations, so that if one frame fits well with a situation, another frame covering a subsequent event would already be made ready for activation. Besides mechanisms for handling defaults, the earlier frame-based representation systems also supported procedural attachment, so that procedures like image or sound processing or robotic action in the real world could be seamlessly integrated.

6 Conclusions

The paper argued that human-centric AI, with its implications of explainability, transparency, robustness, etc., is only going to be possible when AI comes to grips with meaning and understanding. This requires that we go beyond the numerical AI paradigm that is currently dominating AI, where meaning is captured only very indirectly in embeddings and operations over embeddings, but also beyond the symbolic paradigm, which focuses on formal operations over non-grounded symbols.

First of all we need at the very least a form of integrated or hybrid AI that combines numerical and symbolic AI. But we need to go beyond both. The paper argued that a central characteristic of understanding is the ability to build a coherent narrative of an experience based on narratives of past experiences stored in a Personal Dynamic Memory, and integrate this narrative in memory. The big challenge for AI is partly technical, to solve problems of computational complexity to handle the very large knowledge bases and huge inferences that are required. But it is also conceptual. We need to understand much better how new experiences and the narratives built for them get integrated into a Personal Dynamic Memory in such a way that they get triggered again on the most relevant new experiences, and how facts or narratives that are deemed no longer relevant can be forgotten or simply not stored in the first place.

Acknowledgement The author is funded by the Catalan Institute for Advanced Studies (ICREA) embedded in the Institute for Evolutionary Biology (UPF/CSIC) in Barcelona. This work was made possible by H2020 grants within the frame of the Humane AI Flagship preparation project and the AI4EU project.

REFERENCES

[1] L. Steels, ed., 'Computational issues in Fluid Construction Grammar.', **7240**, (2012).

[2] A. Aamodt and E. Plaza, 'Case-based reasoning: Foundational issues, methodological variations, and system approaches.', *AI Communications*, **7(1)**, 39–52, (1994).

[3] L. Barsalou, 'Grounded cognition.', *Annual Review of Psychology*, **59**, 617–645, (2008).

[4] M. Beetz, M. Tenorth, and J. Winkler, *Open-EASE*, 1983–1990, 2015 IEEE International Conference on Robotics and Automation (ICRA), IEEE., 2015.

[5] B. Bergen, N. Chang, and S. Narayan, *Simulated action in an embodied construction grammar*, 108–113, Proceedings of the 26th Annual Meeting of the Cognitive Science Society, L. Erlbaum, Mahwah, NJ, 2004.

[6] J. Bruner, 'The narrative construction of reality.', *Critical Inquiry*, **18(1)**, 1–21, (1991).

[7] V. Dignum, *Responsible Artificial Intelligence. How to develop and use AI in a responsible way.*, Springer International Publishing, 2019.

[8] C. Fillmore, *Frame Semantics.*, volume 10, 111–137, 1982.

[9] Pablo Gervás, Eugenio Concepción, Carlos León, Gonzalo Méndez, and Pablo Delatorre, 'The long path to narrative generation', *IBM Journal of Research & Development*, **63**(1), 1–8, (01/2019 2019).

[10] E. Goffman, *Frame Analysis. An Essay on the Organization of Experience.*, Penguin Books., Harmondsworth, 1974.

[11] A. Goldberg, *Constructions at Work: The Nature of Generalization in Language.*, Oxford University Press., Oxford, 2006.

[12] M. Mateas and P. Sengers (eds.), *Narrative Intelligence*, John Benjamins Pub., Amsterdam, 2003.

[13] J. McCarthy, *Programs with common sense*, Symposium on Mechanization of Thought Processes, National Physical Laboratory, Teddington, Eng, 1958.

[14] M. Minsky, *Semantic information processing*, The MIT Press, Cambridge MA, 1969.

[15] M. Minsky, *A framework of representing Knowledge.*, The Psychology of Computer Vision, McGraw-Hill, New York, 1975.

[16] J. Moore and W. Swartout., *Explanation in Expert Systems: A survey*, ISI Research Reports, ISI, Marina del Rey, Cal, 1988.

[17] A. Nowak, P. Lukowicz, and P. Horodecki, 'Assessing artificial intelligence for humanity: Will AI be the our biggest ever advance? or the biggest threat', *IEEE Technology and Society Magazine*, **37(4)**, 26–34, (2018).

[18] E. Panofsky, *Studies in Iconology. Humanistic themes in the art of the Renaissance.*, Oxford University Press, Oxford, 1939,1972.

[19] J. Pfau, R. Porzel, M. Pomarlan, V. Cangalovic, S. Grundpan, S. Hoefner, J. Bateman, and R. Malaka, *Give MEANinGS to Robots with Kitchen Clash: A VR Human Computation Serious Game for World Knowledge Accumulation*, 85–96, Entertainment Computing and Serious Games, First IFIP TC 14 Joint International Conference, ICEC-JCSG, IFIP, New York, 2019.

[20] Mark O. Riedl, 'Computational narrative intelligence: A human-centered goal for artificial intelligence', *CoRR*, **abs/1602.06484**, (2016).

[21] D. Rumelhart, *Schemata: the building blocks of cognition.*, Theoretical Issues in Reading Comprehension, Lawrence Erlbaum, Hilssdale, New Jersey, 1980.

[22] R. Schank, *Dynamic memory: A theory of reminding and learning in computers and people*, Cambridge University Press, Cambridge Eng, 1990.

[23] R. Schank and Abelson, *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures.*, L. Erlbaum, Hillsdale, NJ, 1977.

[24] C. Shannon, 'A mathematical theory of communication', *SIAM Journal of Scientific and Statistical Computing*, **6**, 865–881, (1985).

[25] D. Sperber and D. Wilson, *Relevance: Communication and Cognition*, Harvard University Press, Cambridge, MA, 1986.

[26] C. Stanfill and D. Waltz, 'Toward memory-based reasoning', *Communications of the ACM*, **29(12)**, (1986).

[27] L. Steels, *Perceiving the Focal Point of a Painting with AI: Case Studies on works of Luc Tuymans.*, 12th ICAART, Springer Verlag, Berlin, 2020.

[28] L. Steels and T. Belpaeme, 'Coordinating perceptually grounded categories through language. a case study for colour.', *Behavioral and Brain Sciences*, **28(4)**, 469–490, (2005).

[29] L. Steels and K. Beuls(eds.), *Case studies in Fluid Construction Grammar*, John Benjamins Pub., Amsterdam, 2019.

[30] L. Steels and R. Lopez de Mantaras., 'The Barcelona declaration for the proper development and usage of Artificial Intelligence in europe.', *AI Communications*, **31(6)**, 485–494, (2018).

[31] L. Steels and M. Hild, *Language grounding in robots*, Springer Verlag, New York, 2012.

[32] L. Talmy, *Toward a cognitive semantics.*, MIT Press, Cambridge Ma, 2000.

[33] E. Tulving, 'Precis of elements of episodic memory.', *Behavioral and Brain Sciences*, **7**, 223–268, (1984).

[34] J. Urbani, S. Kotoulas, J.Massen, F.van Harmelen, and H. Bal, 'Webpie: A web-scale parallel inference engine using mapreduce.', *First Look. Journal of Web Semantics*, **10**, 59–75, (2012).

[35] O. Vilarroya, *Somos lo que nos contamos. Cómo los relatos construyen el mundo en que vivimos.*, Editorial Ariel, Barcelona, 2019.

[36] U. Von der Leyen and et al., 'White paper on artificial intelligence', *EU Commission reports*, (2020).

[37] P. Winston, *The Strong Story Hypothesis and the DirectedPerception Hypothesis*, 2011 AAAI Fall symposium, AAAI Press., Menlo Park Ca, 2011.

[38] W. Xu, 'Toward human-centered AI: A perspective from human-computer interaction..', *IX Interactions*, **26(4)**, 42–46, (2019).