e-ISSN 2723-9640

JoLMA
Vol. 5 — Special issue — October 2024

Wittgenstein, Contexts, and Artificial Intelligence
An Engineer Among Philosophers, a Philosopher Among Engineers

Carlo Penco
Università di Genova, Italy

Abstract Could we take Wittgenstein's philosophy as antagonistic or compatible with AI? Interpretations go in opposite directions. In this paper, I stand with compatibilists and claim that Wittgenstein's discussion on contexts has deep connections with the early stages of AI at different levels. Furthermore, his remarks on context aids in the comprehension of the recent advancement in machine learning based AI, although they embed a warning against the oversimplified association of artificial and human intelligence.

Keywords Wittgenstein. AI. Context. Concept. Family resemblance.

Edizioni Ca'Foscari

Submitted 2024-02-12
Published 2024-03-11

Citation Penco, Carlo (2024). "Wittgenstein, Contexts, and Artificial Intelligence". *JoLMA*, 5, Special issue, 275-294.

DOI 10.30687/Jolma/2723-9640/2024/03/014

1 Introduction

"But machines can't think!" Wittgenstein considers this assertion with doubt, given that he follows remarking that we say "only of a human being and what is like one that it thinks", including dolls and ghosts (PI, § 360). His discussion on the idea that machines can think began in *The Blue Book* and went through the *Remarks on the Foundations of Mathematics* and the *Philosophical Investigations*, and it has often been interpreted as a criticism of Alan Turing, who followed his lectures on the foundation of mathematics in 1939. At the beginning, Wittgenstein considered attributing thinking to machines a category mistake, like attributing colour to numbers (BB, 47) or speaking of "artificial pain" (PG, 64). However, his later remarks are more ambivalent, and interpretations are divided between antagonists, who interpret Wittgenstein's work as a means to contrast AI, and compatibilists, who see Wittgenstein as an inspiration for AI.

Among the antagonists, Dreyfus (1972; 1992) links Wittgenstein's language games to Heidegger's *Dasein* and uses Wittgenstein's conception of "form of life" for denouncing the limits of AI: a form of life cannot be programmed. Harre (1988) relies on the fact that human skills and practices are strictly linked to the human world, while Casey (1988) assumes that we may find in Wittgenstein an argument against the idea that machines can think. However, he claims that the supposed Wittgenstein's argument hides a missing premise that would make the argument invalid or weaker than it appears.[1] Shanker (1998) links Wittgenstein's antagonism towards AI to his criticism of both a mechanistic and psychologistic view of mind where there would be no space for a normative conception of calculation. Although he finds "obscure" Wittgenstein's quotation of Turing machines as "humans who calculate" (RPPI, § 1096), he eventually interprets it as suggesting the relevance of the difference between mechanically following a rule and following a mechanical rule. Fuchs (2022, fn. 8), quoting from Wittgenstein's *Zettel*, insists that a precondition for human interaction is attributing subjectivity to our counterpart, so that we may have shared feelings.

Compatibilists try to bypass Searle's criticism that AI is syntactic

1 Casey presents a reconstruction of Wittgenstein's argument as follows: (i) we are entitled to predicate 'thinking' only to human beings and to what is sufficiently like a human being; (ii) machines are not sufficiently like human beings; (ii) therefore we cannot predicate 'thinking' of machines. He claims that the second premise, to grant the conclusion, should be modal: "Machines cannot be and can never be sufficiently like human beings". But it is difficult to grant that *a priori*. Therefore, it would become empirical and open to debate, given the vagueness of the expression "sufficiently like human beings". See the more complex view by Beran 2014.

and not semantic and lacks intentionality,[2] relying on Wittgenstein's conception of thinking. Obermeier (1983) suggests connecting the concept of thinking to the idea of understanding-as-performance; against Dreyfus, Neumayer (1986) suggests that our interaction with robots may bring about embedding robots in our form of life,[3] and is implicitly followed by Sunday Grève (2023), who speaks of "artificial forms of life". Penco (2012) relies on the idea of understanding as symbol manipulation in context, and Floyd (2019) analyses the connections and philosophical consonance between Turing and Wittgenstein, detailing their meetings and reciprocal influence. Xu (2016) analyses some of the main arguments by Shanker and insists that Shanker's criticism concerns ideas of the old artificial intelligence and does not touch recent connectionist systems that are far away from adherence to psychologism. Molino and Tagliabue (2023) claim that Wittgenstein inspired artificial intelligence, shortly referring to Margaret Masterman, to whose work Wilks (2005) dedicated a careful analysis, on which we devote a section in this paper. Gaver (2023) and Gomes and Selman (2023) enthusiastically put Wittgenstein as ideally inspiring machine learning systems and Large Language Models like ChatGPT.

Who is right? On the side of compatibilism, we have Wittgenstein's attack on the mentalist view of thinking as a hidden process accompanying speech. Against the mentalist view, Wittgenstein defined thinking as "the activity of operating with signs" (BB, 6) or "operating with symbols" (PG, 65), specifying that "'thinking' is a fluid concept, and what 'operating with symbols' is, must be looked at separately in each individual case" (RPPII, §§ 7-8). In face of the question of whether the human body is to be called a machine, he answers, "[i]t surely comes as close as possible to being such a machine" (PI, § 359). These statements are not too far from Newell and Simon's Physical Symbol System Hypothesis (1976), which holds that a physical symbol system has both the necessary and sufficient means for intelligence and that a human is a physical symbol system (as a computer is). Following this analogy, we may interpret his curios remark about Turing machines: "These machines are humans who calculate" (RPPI, § 1096). Differently from Shanker (1998), who devotes quite a bit of space to commenting this passage, it seems to me that this assertion suggests that Wittgenstein was somehow approving Turing's presentation of his machines as "a human being operating with a table of

2 In his paper on the Chinese room, Searle never quotes Wittgenstein, although he can be considered a 'wittgensteinian' for his work on the idea of background.

3 He relies on a quotation where Wittgenstein suggests a possible "language game in which I produce information automatically, information which can be treated by other people quite as they treat non-automatic information" (RPPI, § 817).

rules according to a certain routine" (Floyd 2019, 280-1). Following Floyd, we may remark that Turing used the notion of a human calculator to ground the foundation of logic on socially shareable procedures, and therefore far away from a psychological account or relying on what happens in the mind, in a very Wittgensteinian mood.[4] At the same time, Wittgenstein gives a very open attitude to what we mean by "calculating", saying, just before the remark on Turing machines:

> That we calculate with some concepts and with others do not, merely shews how different in kind conceptual tools are (how little reason we have ever to assume uniformity here). (RPPI, § 1095)

Wittgenstein died in 1951, and it is not awkward to consider the kind of early work in AI, which emerged in the 1950s and developed in the second half of the twentieth century, as part of the spirit of the time. We cannot forget that the first project of AI was presented at Dartmouth College in 1956, five years after Wittgenstein's death, and that part of Wittgenstein's background consisted of studies in engineering.[5] Although compatibilists have made some suggestions regarding these affinities, there is a blind spot in all attempts to propose a Wittgensteinian view of AI: the variety of his remarks on context. I think that Wittgenstein's complex notion of context is what may clarify the deep connection between Wittgenstein's philosophy and the work of AI, and, at the same time, present a warning to a too easy assimilation between artificial and human intelligence.

In what follows, I give an overview of Wittgenstein's remarks on context and their connection with the concepts of family resemblance and language games (§ 2). I then suggest three kinds of analogies or influences of Wittgenstein's ideas on some aspects of the first strand of AI and on the beginning of information retrieval, which ideally extends to more recent results (§ 3). Eventually, I will discuss his notion of learning and understanding to gain a better grasp on the difference between artificial intelligence and human understanding (§ 4).

4 Shanker insists on the difference between the first works of Turing and the Turing's works after the fourties, and takes a unitarian view of Wittgenstein's remarks, while it seems that Wittgenstein slowly changed his mind from the early thirties to the later years, after the publication of Turing (1936).

5 We may see traces of his studies in engineering in his continuous reference to technical problems, even in the presentation of classical philosophical theories, like Frege's context principle, as in the following remarks: "If we say: A word only has meaning in the context of a proposition (*satzzusammenhang*), then that means that it's only in a proposition that it functions as a word, and this is no more something that can be said than […] that a cogwheel only functions as such when engaged with other cogs" (PR, 12). Or also see: "A word only has meaning in the context of a proposition (*Satzbervand*): that is like saying only in use is a rod a lever. Only the application makes it into a lever" (PR, 14).

2 Wittgenstein's Contexts

The first occurrences of the term "context" concern Frege's contextual principle presented in the *Grundlagen*: never to ask for the meaning of a word in isolation, but only in the context of a sentence.[6] Eva Picardi (2010) analyses the difference between Frege's and Wittgenstein's context principles. First, Wittgenstein rejects the Fregean idea that sentences can be considered names whose reference (*Bedeutung*) is a truth value. For Wittgenstein, names have reference but not sense, while sentences have sense (truth conditions) but no reference. Second, sentences express states of affairs, and this feature opens a new element in a possible semantic framework (something analogous to what later came under the tag of 'aboutness'). Although this aspect may be rediscovered when discussing the early Frege, as Perry (2019) did, the reference to states of affairs, situations, or circumstances marks a further departure from Frege's framework. Third, after 1930, Wittgenstein takes another step forward, widening Frege's principle: besides speaking of the meaning of a word as dependent on the context of a sentence, he begins to speak of the meaning of a sentence as depending on the context in which it is uttered. Here we may think of the meaning of a sentence, à la Robert Brandom, as inferential potential, or as a set of presuppositions, entitlements, commitments, inferences, or conversational implicatures. All of these elements may find clarification when the sentence, as Wittgenstein suggests, is understood as uttered in the context of a play, a theatrical performance, or a drama (LWI, § 38).

However, as Picardi (2005) remarks, we should be careful not to widen this generalisation of context to a holistic view of meaning, as done by Davidson and Brandom. Wittgenstein was very careful to keep the idea of contextual dependence always delimited to specific language games and on the idea that there are rules constitutive of some concepts (for instance, defining logical constants). Concerning his discussion on contexts, Wittgenstein's thought is not completely at home neither with contextualism nor with holism. I do not

6 Frege: "Nach der Bedeutung der Wörter muss im Satzzusammenhange, nicht in ihrer Vereinzelung gefragt warden". Wittgenstein: "Nur im Zusammenhang des Satzes hat ein Name Bedeutung" (TLP, 3.3); "Nur im Satzzusammenhang hat ein Wort Bedeutung" (PR, 12); "Ein Wort hat nur in Satzverband Bedeutung" (PR, 14). Besides the relation between words and sentences, we also have the corresponding relation between objects (the references of names) and states of affairs: "Wenn ich mir den Gegenstand im Verbande des Sachverhalts denken kann, kann ich ihn nicht außerhalb der M ö g l c h k e i t dieses Verbandes denken" (TLP, 2.0121). "If I can imagine objects combined in states of affairs, I cannot imagine them excluded from the possibility of this context." The translation from Pears and McGuinness speaks of "possibility of such combination" and I changed coherently with Wittgenstein's suggestion to Ogden to translate the passage with the term "context".

enter here in the discussion on semantic holism, to which I devoted some space elsewhere.[7] I only remark that there is a problem with the general agreement of the identification of Wittgenstein's ideas with contextualism, as represented by Robyn Carston, Francois Recanati, Charles Travis, and many others who centre their core ideas on the notion of underdetermination of meaning. The problem is that, besides the idea of a deep grammar characterising kinds of language games, Wittgenstein speaks of "ordinary context" and "ordinary sense" as if he accepted the notion of "literal meaning" in contrast with extreme contextualism. Wittgenstein defines "ordinary context" in an interesting discussion of sentences where we have no criteria of application, suggesting that ordinary meaning is linked to standard criteria of application.[8] He also speaks of the use of a word in the context of the language game "which is its original home" (PI, § 116). He thereby suggests the concept of stereotypical meaning or prototype, a term he uses in *The Blue Book* and a subject that Hilary Putnam, Eleanore Rosch, and Marvin Minsky have all developed in various ways.[9] Surely, Wittgenstein has been a source of inspiration to many, but it is difficult to compare his remarks on context with contemporary philosophy of language, just as it is difficult to compare a "sketch of landscapes" (PI, preface) with detailed and alternative maps of the same landscape. Precisely defined concepts like Kaplan's context of utterance, Stalnaker's context set, or Recanati's context-sensitivity are not to be found in Wittgenstein's works, although there are hints towards some of those directions of research (like, for instance, the idea of the meaning of a sentence as an "expansion" (OC, § 349)).

7 Robert Brandom follows Davidson's acceptance of Quine's holism, which was supported in *Word and Object* by *The Blue Book*'s quotation that understanding a sentence is understanding a language. However, Wittgenstein speaks of "a" language, and we may interpret Wittgenstein's view as "understanding a sentence is understanding the language game in which it makes sense" (BB, 5), which amounts to a form of local holism, or weak molecularism, as Michael Dummett insisted. *Global* holism (or even a strong form of molecularism that leads to holism) claims that if two people share a belief p, there is some other belief q, which must also be shared: ∀ p ∃ q (q≠p & Nec (p is shared → q is shared). On the contrary, local holism would account for something weaker: necessarily, if you share p, there are other beliefs that are also shared. However, there is no privileged set of beliefs that must be shared; it is only necessary that, if p is shared, some not previously determined belief should be shared: ∀ p Nec (p is shared → ∃ q (q≠p & q is shared). See Perry 1994 and Penco 2001; 2004.

8 See BB, 10. Besides speaking of "ordinary language" and "ordinary context", the idea of an *ordinary* meaning or ordinary use of a word is a topic that often recurs in his works, especially in BB (27, 36, 52, 53, 62, 66, 140), but also in PI (§§ 258, 344, 351, 418, 536, 615, and PI II, 176, 192) and in RPPI (§§ 52, 99, 126, 358) where we find also the idea of "ordinary language game" (§ 820).

9 On the difference between stereotype and prototype, see Marconi 1997, 22-8, who discusses the relationships between the ideas by Rosh, Putnam, and Minsky.

However, there is more to say about Wittgenstein's different applications of the term "context", a term that Wittgenstein often uses but never mentions or thematises as such, implicitly suggesting that it might be a perfect case of family resemblance predicate.[10] Beside the Fregean context principle (TLP, 3.3; PR, 12, 14; PI, § 50), he uses the term "context" in different ways: as spatial context (PG, 88; PI, § 539; RC, § 255) or context of perception (RPPI, § 531; PI, II, xi, concerning the duck-rabbit figure), context of conversation or context of speech (PG, 79; Z, § 311; LWI, §§ 118-20; OC, § 349), and eventually context *as circumstance* or the set of circumstances in which a person speaks (PG, 28, 88; PI, §§ 203, 539; RFM, V, § 45; RPPI, § 331; LWI, §§ 253-4; OC, § 662). To understand a sentence, you have to look at the context, intended as the circumstances or situations in which people interact (for instance, as already mentioned, the situation described in a play).

Context as a situation or set of circumstances is a concept that has been developed in different ways in philosophy and computer science as well, from Barwise and Perry's situational semantics to Margaret Masterman's theories of semantic classification for information retrieval and John McCarthy's multi-context theory. I was surprised to verify that there were no explicit connections between situational semantics and multi-context theory, although Barwise, Perry, and McCarthy shared a common ground at the University of Stanford. Often philosophers and computer scientists work on parallel lines and therefore do not immediately converge.[11] In this historical paper, I cannot fill the gap but only show some connections between Wittgenstein's work and the computer scientists' work.

3 Connections Between Wittgenstein's Ideas of Context and AI

We can devise three main lines of the connection between Wittgenstein's ideas of context and early AI: (i) the idea of family resemblance insofar as it derives from the idea of privileged or ordinary contexts on which we rely and from which we may define similar ones by analogy; (ii) the idea of language games insofar as it derives from the idea

10 In a different setting, McCarthy, Buvač (1994, 45) claim that "the term 'context' will appear in useful axioms and other sentences but will not have a definition involving 'if and only if'".

11 However, even computer scientists do not interact enough. Shoham (1991, 395) wondered about the interest "to examine the extent to which work on situation semantics can be usefully applied in AI". Just a few years before, Terry Winograd (1985) had written a paper on the possible use of situation semantics for the development of more expressive programming languages. There were no connections between the two ideas.

that the meaning of words and sentences depends on particular delimited situations or circumstances in which we may better understand how language works; (iii) the criticism of mathematical logic and formal semantics prompting the first use of Wittgenstein's ideas for information retrieval.

(i) Family resemblance, concepts and contexts

Among the first representatives of artificial intelligence, we find John McCarthy and Marvin Minsky, who, besides apparent differences, worked on the same kinds of problems: how can we treat reasoning when new information comes and compels us to change our premises? How do we treat language processing when our concepts are vague and cannot be defined by necessary and sufficient conditions? A chair is a four-legged piece of furniture, but it may have three legs or even one. Birds fly, but some birds do not: penguins, embalmed birds, and birds with broken wings. Minsky answered with the idea of concepts as frames with default values (a chair has 4 legs or birds fly unless some contradictory situation triggers a change in these values). Minsky (1974) explicitly used Wittgenstein's idea of family resemblances to introduce his idea of frames as stereotyped situations with default information. If everything goes, the standard frame is accepted by default, but we may find unexpected differences and therefore we should have features ('demons') that suggest what to do if the expectations of the frame are not fulfilled (for instance, if a chair does not have four legs). Therefore, the default values of a frame can change depending on context, and frames will be connected with other frames, like an enriched semantic network. Minsky used the Wittgensteinian idea of a "network of overlapping and crisscross resemblances" as a way to explain

> how we can feel as though we know what a chair or a game is—yet we cannot always define it in a 'logical' way as an element in some class hierarchy or by any other kind of compact, formal, declarative rule. (Minsky 1974, 51)

While Minsky was looking for some alternative with respect of mathematical logic, McCarthy – who could be called the grandfather of the great old-fashioned AI (GOFAI) – tried to make mathematical logic more adaptable to the vagaries of commonsense reasoning. McCarthy (1986) used non-monotonic logic (circumscription) to accept changes in the conclusions when some abnormality enters the set of information (all 'normal' birds fly, but we can change this property in front of an 'abnormal' bird).

A similarity between McCarty's multi-context theory and Wittgenstein's view also concerns the idea of the impossibility of a complete description: if I want to take a flight, I just need to know the timetable and buy the ticket. But if I lose the ticket, I need to describe

the situation with more details on how to recover the ticket, and if the flight is cancelled, I need to enrich my description of the situation with more information. There is no unique context in which I may have an absolute complete description. The completeness of the description depends on the needs of the particular circumstance in which I am in and on underlying assumptions. Analogously, probably referring to the standard problems of Russell's theory of description, Wittgenstein was very well aware that what belongs to a "complete description" will depend on "the purpose of the description" (RFM, VII, § 311). Referring to the idea of Quine's "eternal sentences", which, according to Quine, do not depend on context, McCarthy remarks that they do not exist and that all sentences are dependent on the kind of context chosen. We may only reach a "relative decontextualization", where some common context is explicitly expressed.

(ii) Contexts as situations: language as a motley of language games

We find the most striking similarity with Wittgenstein's ideas in McCarthy's multi-context theory, where contexts are defined as theories with their axioms, their domain, and their rules, creating a vision of language as an unordered series of different local theories that could be considered a good approximation of the idea of different language games, where no sentence can have meaning "out of any context", but can always be considered as having a meaning in particular contexts or particular circumstances (PI, II, ii; OC, §§ 349, 532, 553, 662; RPPI, § 1037).

One of the first computer science exemplifications of a language game is Winograd's SHRDLU, a program for a dialogue with a simulated robot with which to interact in a toy world of boxes, cubes, and pyramids of different forms and colours. The game was very simple: giving orders on how to move those blocks, asking questions, or giving names for new arrangements of those blocks. There is a striking similarity between this toy world and Wittgenstein's builders.[12]

Where are the similarities? Both Wittgenstein and Winograd realised that they could give a good analysis of the workings of language if they considered simplified situations: the knowledge of the toy world permits simple linguistic interactions, similar to the interaction of the builders. Besides, both examples are an expression of the idea of language as a kind of action in context. While Wittgenstein was developing his view of language games as a mixture of language and actions, Winograd (1972) was using Austin's classification of basic speech acts (question, command, assertion) in his

12 I was impressed by listening to Winograd's presentation of SCHRDLU in 1972 during a Pisa conference organised by Antonio Zampolli, chief of the laboratory of computational linguistics where I was working at making punched cards. Some years later, I presented the comparison between Winograd's SHRDLU and Wittgenstein's builders with Marcello Frixione at a meeting of EECSE in Camogli (Italy) in 1994.

interaction with the simulated robot, in a view of language "as action rather than structure or the result of a cognitive process" (Winograd 1980, 230). Every speech act triggers a procedure: a command triggers an operation on the blocks, a question the giving of information, an assertion the storing of information. The meaning of a word is represented as a procedure that permits finding the referent of the expressions in the sentence (in our case, the objects are simple and composed blocks). The basic idea underlying Wittgenstein's and Winograd's views of language was similar: simplified models of language interaction suggest the idea of language as a heterogeneous set of diverse toy worlds or various language games or, in McCarthy's terminology, of various contexts. The difference concerns the aims: computer scientists aim to provide a good representation of natural language processing (NLP) for creating working programs, while Wittgenstein's philosophical aim is to remove specific misunderstandings, and even his criticisms towards Turing concern not his mathematical theory, but the possible psychologistic interpretation of the idea of thinking machines.

The perspective of multi-context theories maps even more strictly the attitude of Wittgenstein towards language as a mixture of language games where there is no meta-language game over the others – Wittgenstein speaks of mathematics as "a motley of techniques of proofs" (RFM, III, § 46). Analogously, for McCarthy, there is no universal context but just different contexts with their own rules. In multi-context theory, we find something new with respect to the role of contexts in describing the workings of language: not only words and sentences, but also rules depend on the context we are considering. For McCarthy and Buvač (1994), some rules may be common to different contexts, permitting us to let what has been derived in one context enter another context. In this way, McCarthy introduced a new problem: how to individuate the relations among contexts, and this problem maps Wittgenstein's idea of intermediate members of different language games (see Penco 2004; 2007), or his requirement to look for a wider context to change interpretation of what is said (PG, 88; PI, §§ 539, 686; RPPI, § 1066).

The effort of McCarthy was to show which rules may govern relations among contexts, how to find whether contexts are compatible with one another or not, how the conclusions reached inside a context may be valid in another context or not, how we can change a context either by enriching it or simplifying it, or by making parameters explicit or leaving them implicit, how we can leave a context to enter another with different rules, and so on.[13] What emerges from these

13 On these kinds of rules see Benerecetti, Bouquet, Ghidini 2000, quoted by Guha, McCarthy 2003.

early works is a representation of language as a non-ordered mixture of different contexts, or, if you like, of different toy worlds or different language games related one another by different rules without a general universal structure underlying all of them.

(iii) Criticism of mathematical logic and information retrieval

Margaret Masterman (MM) was one of the five students to whom Wittgenstein dictated *The Blue Book* in 1933-34.[14] Later, in 1955, she founded the Cambridge Language Research Unit (CLRU), where she promoted a view of natural language processing against both the semantic analysis of Richard Montague and the syntactic analysis of Noam Chomsky. Her view was relying on the use of thesauri to find patterns or underlying structures of semantic relations. For MM (2005, 109), a thesaurus is a "language system classified as a set of contexts", where contexts are sentences in which a word appears, following the first traditional definition of meaning as the use of a word in a sentence, but also defined by semantic clusters. A few aspects of her research reveal the influence of Wittgenstein:

– MM gave great relevance to the concept of context as a kind of situation: she claims that, notwithstanding differences in language and culture, we "can share a common stock of extra-linguistic contexts" (2005, 127), which is a common stock of kinds of situations. As example of lack of shared extra-linguistic contexts she presents the comparison between the forms of life of humans and ants.[15]

– MM maintained a distance towards formal logic as an analysis of natural language:

> [F]ormal logic as we at present have it is not and cannot be directly relevant to the contextually based study of semantic pattern. Logic is the study of relations, and, in particular, it is the study of derivability. (2005, 261)

Semantic patterns are not related to mathematical logic, but are derived from the contexts in which a word appears:

> [T]he use of a word is its whole field of meaning, its total 'spread'. Its usages, or main meanings in its most frequently found contexts, together make up its Use. (2005, 126)

14 On the relation between Margaret Masterman and Wittgenstein see Wilks 2005; 2007 and Liu 2021. Among many others, Margaret Masterman deeply influenced Yorik Wilks, Margaret Boden and Kwame Anthony Appiah.

15 Humans sleep and dream, while the latter do not, and this makes it impossible to share a common stock of situations. It is curios that Wittgenstein used an example about ants in showing the difference between humans' natural history and other species (RPPII, §§ 22-4).

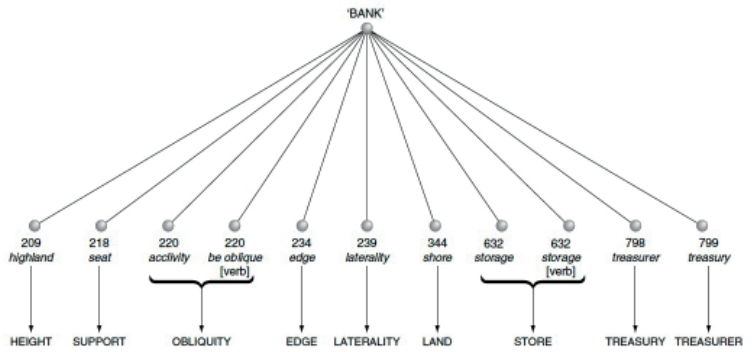

Figure 1 Scheme from Masterman 2005, 288

– Following Wittgenstein's example of the ambiguity of the term "bank" (BB, 11-13), MM presented the term "bank" together with the different contexts or "subjects-which-it-is-most-used-to-talk-about" (2005, 289), speaking of those different subjects as "quasi-Wittgensteinian families", suggesting a comparison with Wittgenstein's idea of family resemblance predicates **[fig. 1]**.

– Beyond the analysis of the same words used to refer to different concepts (like "bank") or to be embedded in different patterns, MM discusses the problem of different expressions used to refer to the same situation. The basic idea is that there is no proper complete synonymy – if not, it would be useless – but what different people say about the same kinds of situations largely overlaps and may form clusters of meaning and also clusters of overlapping contexts (2005, 69).

Masterman's strategy to find similarities, clusters, and patterns in a thesaurus will become central to the CRLU information retrieval system. Works in the tradition of Margaret Masterman used strategies of letting patterns or clusters of words emerge from unsupervised statistical methods on a large data set, following Wittgenstein's idea of family resemblance.[16] Those works, as Halpin (2011) remarks, were at the source of the first Web search engines. Referring to the

16 See for instance Karen Spärck Jones (1986, 64), who quotes Wittgenstein's method of similarities and differences concerning her view of treating synonymy. She relied on Roger Needham's statistical theory of clumps, a particular explication of the idea of family resemblances, for which words can be defined in terms of statistical clumps of other words. See also Wilks 2008, who connects Wittgenstein's ideas to his preferential semantics, Shank's conceptual dependency, or Fillmore's case grammar.

contrast between the use theory and the causal theory of meaning, Halpin (2011, 18) remarks that "search engines like Google embody an alternative theory of meaning, one based on an objective notion of sense implicitly given by Wittgenstein". In a way, the inventors of Google, Sergey Brin and Larry Page, still working with his tutor Terry Winograd,[17] developed a search engine whose basic ideas come from a long history, ideally reconnecting them to Wittgenstein's influence. Margaret Masterman, missing the technical means we have today, was probably very much in advance of her times and anticipated some of the most recent trends in extracting patterns from learning algorithms. Wilks (2008) and Molino and Tagliabue (2023), referring to Masterman's work, give more detailed comments on the development of machine learning algorithms following this kind of trend up to the more recent LLMs.

4 Understanding and Learning Language in Context

Having presented the affinity of Wittgenstein's ideas of context dependence with some features of artificial intelligence projects, we may come back to the general question: do machines think? Following Tarski,[18] we might say: "Of course they can, it only depends on what you mean by 'think'". If we equate thinking as "operating with signs" and understanding language with "to be master of a technique" (PI, § 199), it would be very difficult to avoid the conclusion that contemporary chatbots or advanced robots master a technique of language use and therefore think. The problem is that the technique of "thinking machines" is different from the technique used by humans, just as the technique of flying aeroplanes is very different from the technique of flying used by birds. Therefore, assuming that understanding a language means mastering a technique and that machines 'understand language' (in the limited sense that they pass the Turing test), where does the difference lie between human understanding and machine understanding?

AI based on deep learning algorithms is very different from the workings of the human mind, as strongly remarked by Chomsky, Roberts and Watumull (2023). Chomsky's point concerns not only the concept of understanding but also the concept of learning. Learning is a family resemblance concept, and Wittgenstein gave examples of different ways we use this concept:[19] learning a language (PI, § 7),

17 Winograd himself contributed to the presentation of the page-rank algorithm. See Page et al. 1998.

18 Quoted in Obermeier 1983, 347.

19 See Williams 1999 and Vazquez Hernandez 2020.

learning the numerals by heart and learning to use them (PI, § 9), learning an unknown language from ostensive definitions (PI, § 32), and different ways of learning the meaning of a word, like the difference between "to point to this thing" and "to point to the colour, not to the shape" (PI, § 35), or the different examples and different language games with which we learn the word "good" (PI, § 77), and so on.[20] One may therefore think that we may see the technique of machine learning that governs most LLMs as a new kind of learning procedure, a realisation of the idea of Margaret Masterman of extracting semantic patterns from data. Learning algorithms, after training, are let loose to autonomously find further patterns inside texts. Could this be compared to human learning? Certainly it can, but the difference is so huge that our concepts of thinking and understanding that depend on the way we learn language are stretched *near* a breaking point. Let us see the differences.

The first difference is quantity: humans learn to speak with a very small amount of linguistic data; this means that they learn language in an extralinguistic context, and rely on some kind of innate system (brought about by evolution) to master the technique of linguistic interaction. Learning algorithms learn from a huge amount of data, both visual patterns and linguistic content taken from many already organised data sets, and therefore produce a second order intelligence grounded on our examples of natural language, which brings us to the second difference.

The second difference is quality: AI dialogue systems take their content from big data sets, of technical or literary content. Therefore they begin with the highest result of our civilisation (included programming languages or social networks), including biases depending on the data sets used. On the contrary we begin to learn with more emotional and basic stuff, mostly linked to perception, emotion, and physical interaction. Our body, emotions, and the context of perception, have an essential part in learning language, and only after many years we humans begin to reach the abstract concepts and connections that are given to the machines since the outset.

The third difference is classification: the clusters and patterns arrived at by machine learning algorithms are not necessarily similar to the way we classify concepts and situations. Logicians like Frege began to find patterns in sentences that permit us to better understand our logical reasoning, but the patterns individuated by new learning

20 I just used some examples from the first pages of the *Philosophical Investigations*, but the discussion on different aspects of 'learning' goes through all his published books, where the word 'learn' in different ways is reported in almost 400 quotations. The reference to the word 'good' is reminiscent of Wittgenstein's reaction to Moore's lessons on the term 'good', where he found one of the first examples of family resemblance concepts (see Vaccarezza, Penco 2023).

algorithms are more linked to clusters identified in a way even programmers cannot know. Although those algorithms may have some procedure similar to ours, the patterns they use are such that we cannot grasp them in the same way we grasp our concepts.

But what does it mean 'to grasp' a concept or a thought? A semantic theory may claim that mastering a thought is mastering its inferential potential, the commitments undertaken in asserting the corresponding sentence. When we speak of commitments, we introduce an ethical issue and not only a cognitive issue. The thought experiments of a drone that kills the operator in order to perform its task[21] introduces a new vision of what it means to grasp a thought: you need to master the context and the complex circumstances in which you are, and AI lacks this kind of complexity. It is a complexity that cannot be reduced to a technique. We have been impressed by the idea that understanding a language means mastering a technique. However, besides different kinds of techniques for mastering a language, Wittgenstein suggests that there is also a kind of learning that is not given by any technique.

Wittgenstein speaks of the capacity to give 'expert' judgements about the genuineness of expressions of feeling. AI experts have invented programs that detect shifts in speech rate, pitch, volume, or microtremors undetectable by humans to verify the 'real' feeling or honesty of a person. Who knows? Maybe it is even simpler than NLP. But Wittgenstein speaks of learning 'by experience' as a kind of understanding that only experienced people can teach. There are rules linked to this experience, but they are different from 'calculating rules'. This last remark seems to make a distinction among different kinds of understanding, one linked to a technique, the other linked to context understanding that no technique can offer (PI, II, xi, 227; LWI, §§ 917-27).

5 Summary and Conclusion

At the end of the day, Wittgenstein was more open to the idea of thinking machines than it is often claimed, and the compatibilists have some reasons to see Wittgenstein's work as a forerunner of artificial intelligence, starting with his influence on Turing and Masterman and the similarity of his views and those of the early AI experts. The point of Wittgenstein's remarks on contexts is the search for differences: different kinds of learning show that there are and will be different ways of thinking and understanding linked to the different kinds of contexts in which we are in. There is enough space for

21 See the discussion in Davis, Squire 2023.

different contexts and different language games to contain both human and artificial thinking and their interactions. However, a warning remains central: what counts is awareness of the difference between AI ways of reasoning and human ways of reasoning. The latter may contain some specific capacities, resulting from different learning procedures and the expertise that arises from them.[22]

References

Appiah, K.A. (2008). "The Epiphany Philosophers". *New York Times Magazine*, September 21.

Barwise, J.; Perry, J. (1983). *Situations and Attitudes*. Cambridge, MA.; London: MIT Press.

Benerecetti, M.; Bouquet, P.; Ghidini, C. (2001). "On the Dimensions of Context Dependence: Partiality, Approximation, and Perspective". *LNAI*, 2116, 59-72. https://doi.org/10.1007/3-540-44607-9_5.

Beran, O. (2014). "Wittgensteinian Perspectives on the Turing Test". *Studia Philosophica Estonica*, 35-57. https://doi.org/10.12697/spe.2014.7.1.02.

Brandom, R. (1994). *Making It Explicit*. Cambridge, MA.: Harvard University Press.

Casey, G. (1988). "AI and Wittgenstein". *Philosophical Studies*, 32, 156-75. https://doi.org/10.5840/philstudies19883239.

Chomsky, N.; Roberts, I.; Watumull, J. (2023). "The False Promise of ChatGPT". *The New York Times*, March 8.

Davis, C.R.; Squire, P. (2023). "Air Force Colonel Backtracks Over His Warning About How AI Could Go Rogue and Kill Its Human Operators". *Business Insider*. businessinsider.com/ai-powered-drone-tried-killing-its-operator-in-military-simulation-2023-6.

Dreyfus, H.L. (1972). *What Computers Can't Do*. New York: Harper & Row.

Dreyfus, H.L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA.; London: MIT Press.

Floyd, J. (2019). "Wittgenstein and Turing". Mras, G.M.; Weingartner, P.; Ritter, B. (eds), *Philosophy of Logic and Mathematics = Proceedings of the 41st International Ludwig Wittgenstein Symposium*. Berlin; Boston: De Gruyter, 263-96. https://doi.org/10.1515/9783110657883-016.

Fuchs, T. (2022). "Understanding Sophia? On Human Interaction with Artificial Agents". *Phenomenology and the Cognitive Sciences*, 23(2), 1-22. https://doi.org/10.1007/s11097-022-09848-0.

Gaver, M. (2023). "From Wittgenstein's Language Games to Large Language Models". YouTube video on The Times of AI channel.

Gomes, C.; Selman, B. (2023). "Exploring Connections – Wittgenstein, Gödel, and ChatGPT". YouTube video on Vienna Center For Logic and Algorithm channel.

Guha, R.; McCarthy, J. (2003). "Varieties of Contexts". *LNAI*, 2680, 164-77. https://doi.org/10.1007/3-540-44958-2_14.

22 Thanks to Margherita Benzi, Diego Marconi and Giovanni Mion for suggestions on an earlier version of this paper.

Halpin, H. (2011). "Sense and Reference on the Web". *Minds and Machines*, 21, 153-78. https://doi.org/10.1007/s11023-011-9230-6.

Harre, R. (1988). "Wittgenstein and Artificial Intelligence". *Philosophical Psychology*, 1(1), 105-15. https://doi.org/10.1080/09515088808572928.

Liu, L.H. (2021). "Wittgenstein in the Machine". *Critical Inquiry*, 47(3), 425-55. https://doi.org/10.1086/713551.

Marconi, D. (1997). *Lexical Competence*. Cambridge, MA; London: MIT Press.

Masterman, M. (2005). *Language, Cohesion and Form: Selected Papers*. Edited by Y. Wilks. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486609.

McCarthy, J. (1986). "Applications of Circumscription to Formalizing Common-sense Knowledge". *Artificial intelligence*, 28(1), 89-116. https://doi.org/10.1016/0004-3702(86)90032-9.

McCarthy, J. (1990). *Formalizing Common Sense*. Edited by V. Lifschitz. Norwood, NJ: Ablex Publishing Corporation.

McCarthy, J.; Buvač, S. (1994). "Formalizing Context (Expanded Notes)". *Technical Report STAN-CS-TN-94-13*. http://i.stanford.edu/pub/cstr/reports/cs/tn/94/13/CS-TN-94-13.pdf.

Minsky, M. (1974). "A Framework for Representing Knowledge". *MIT-AI Laboratory Memo 306*.

Molino, P.; Tagliabue, J. (2023). "Wittgenstein's Influence on Artificial Intelligence". https://doi.org/10.48550/arXiv.2302.01570.

Neumayer, O. (1986). "A Wittgensteinian View of Artificial Intelligence". Born, R. (ed.), *Artificial Intelligence. The Case Against*. London: Croom Helm, 132-73. https://doi.org/10.4324/9781351141529-6.

Newell, A.; Simon, H.A. (1976). "Computer Science as Empirical Inquiry: Symbols and Search". *Communications of the ACM*, 19(3), 113-26. https://doi.org/10.1145/360018.360022.

Obermeier, K.K. (1983). "Wittgenstein on Language and AI: The Chinese-room Thought Experiment Revisited". *Synthese*, 56(3), 339-49. https://doi.org/10.1007/BF00485470.

Page, L. et al. (1998). "The PageRank Citation Ranking: Bringing Order to the Web". *Technical Report, Stanford Digital Libraries SIDL-WP-1999-0120*, 161-72.

Penco, C. (2001). "Local Holism". *LNAI*, 2116, 290-303. https://doi.org/10.1007/3-540-44607-9_22.

Penco, C. (2004). "Wittgenstein, Locality and Rules". Coliva, A.; Picardi, E. (eds), *Wittgenstein Today*. Padova: Il Poligrafo, 249-74.

Penco, C. (2007). "Idiolect and Context". Auxier, R.E.; Hahn, L.E. (eds), *The Philosophy of Michael Dummett. Library of Living Philosophers*, vol. 31. Chicago and La Salle (IL), Open Court, 567-90.

Penco, C. (2012) "Updating the Turing Test. Wittgenstein, Turing and Symbol Manipulation", *Open Journal of Philosophy*, 2, 189-94. https://doi.org/10.4236/ojpp.2012.23029.

Perry, J. (1994). "Fodor and Lepore on Holism". *Philosophical Studies*, 73, 123-38. https://doi.org/10.1007/BF01207661.

Perry, J. (2019). *Frege's Detour*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198812821.001.0001.

Picardi, E. (2005). "Was Frege a Proto-Inferentialist?". Reprinted in Picardi 2022. https://doi.org/10.1093/oso/9780198862796.003.0014.

Picardi, E. (2010). "Wittgenstein and Frege on Proper Names and the Context Principle". Reprinted in Picardi 2022. https://doi.org/10.1524/dzph.2009.0051.

Picardi, E. (2022). *Frege on Language, Logic and Psychology: Selected Essays*. Edited by A. Coliva. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198862796.001.0001.

Shanker, S. (1998). *Wittgenstein's Remarks on the Foundations of AI*. London; New York: Routledge.

Shoham, Y. (1991). "Varieties of Context". Lifschitz, V. (ed.), *Artificial Intelligence and Mathematical Theory of Computation: Papers in Honor of John McCarthy.*New York: Academy Press, 393-408. https://doi.org/10.1016/B978-0-12-450010-5.50028-1.

Spärck Jones, K. (1986). *Synonymy and Semantic Classification*. Edinburgh: Edinburgh University Press.

Sunday Grève, S. (2023). "Artificial Forms of Life". *Philosophies*, 8(5), 89. https://doi.org/10.3390/philosophies8050089.

Turing, A. (1936-7)."On Computable Numbers, with an Application to the Entscheidungsproblem". *Proceedings of the London Mathematical Society*, 42(2), 230-65. Reprinted in Davis, M. (ed.) (1965), *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*. Hewlett, NY: Dover Publication, 116-54.

Vaccarezza, S.; Penco, C. (2023). "Wittgenstein's Non-Non-Cognitivism". Dreon et al. (a cura di), *Senza Trampoli. Saggi filosofici per Luigi Perissinotto*. Milano; Udine: Mimesis.

Vázquez Hernández, A. (2020). *Wittgenstein and the Concept of Learning in Artificial Intelligence* [MA Dissertation]. Bergen: University of Bergen.

Williams, M. (1999). "The Philosophical Significance of Learning in Wittgenstein's Later Philosophy". Williams, M., *Wittgenstein, Mind and Meaning: Toward a Social Conception of Mind*. London; New York: Routledge, 188-215.

Wilks, Y. (2005). "Editor's Introduction". Masterman 2005, 1-18. https://doi.org/10.1002/9780470776599.ch.

Wilks, Y. (2007). "A Wittgensteinian Computational Linguistics?". *Epiphany Philosophers Blog*. http://epiphanyphilosophers.org/wp-content/uploads/2017/01/wittgenstein.pdf.

Wilks, Y. (2008). "What Would a Wittgensteinian Computational Linguistics Be Like?". *Proceedings of Convention of the Society for the Study of AI and Simulation of Behavior. Workshop on Computers and Philosophy*, 1st-4th April 2008, University of Aberdeen.

Winograd, T. (1972). "Understanding Natural Language". *Cognitive Psychology*, 3(1), 1-191. https://doi.org/10.1016/0010-0285(72)90002-3.

Winograd, T. (1980). "What Does it Mean to Understand Language?". *Cognitive Science*, 4(3), 209-41. https://doi.org/10.1207/s15516709cog0403_1.

Winograd, T. (1985). "Moving the Semantic Fulcrum". *Linguistics and Philosophy*, 91-104. https://doi.org/10.1007/BF00653378.

Wittgenstein, L. (1958). *Philosophical Investigations*. 2nd edition. Edited and transl. by G.E.M. Anscombe. Oxford: Basil Blackwell. [PI]

Wittgenstein, L. (1969). *The Blue and The Brown Books*. 2nd edition. Oxford: Blackwell. [BB]

Wittgenstein, L. (1974a). *On Certainty.* Reprinted with corrections and indices. Edited by G.E.M. Anscombe; G.H. von Wright, transl. by D. Paul; G.E.M. Anscombe. Oxford: Blackwell. [OC]

Wittgenstein, L. (1974b). *Philosophical Grammar.* Edited by R. Rhees, transl. by A. Kenny. Oxford: Blackwell. [PG]
Wittgenstein, L. (1975). *Philosophical Remarks*. Edited from his posthumous writings by R. Rhees, transl. by R. Hargreaves; R. White. Oxford: Blackwell. [PR]
Wittgenstein, L. (1978a). *Remarks on Colour.* Edited by G.E.M. Anscombe, transl. by L.L. McAlister; M. Schättle. Oxford: Blackwell. [RC]
Wittgenstein, L. (1978b). *Remarks on the Foundations of Mathematics*. 3rd edition. Edited by G.H. von Wright; R. Rhees; G.E.M. Anscombe, transl. by G.E.M. Anscombe. Oxford: Blackwell. [RFM]
Wittgenstein, L. (1980a). *Remarks on the Philosophy of Psychology*, vol. 1. Edited by G.E.M. Anscombe; G.H. von Wright, transl. by G.E.M. Anscombe. Oxford: Blackwell. [RPPI]
Wittgenstein, L. (1980b). *Remarks on the Philosophy of Psychology*, vol. 2. Edited by G.H. von Wrightf; H. Nyman, transl. by C.G. Luckhardt; M.A.E. Aue. Oxford: Blackwell. [RPPII]
Wittgenstein, L. (1981). *Zettel*. 2nd edition. Edited by G.E.M. Anscombe; G.H. von Wright, transl. by G.E.M. Anscombe. Oxford: Blackwell. [Z]
Wittgenstein, L. (1982). *Last Writings on the Philosophy of Psychology*, vol. 1. Edited by G.H. von Wright; H. Nyman, transl. by C.G. Luckhardt; M.A.E. Aue. Oxford: Blackwell. [LWI]
Xu, Y. (2016). "Does Wittgenstein Actually Undermine the Foundation of AI?". *Frontiers of Philosophy in China*, 11(1), 3-20.