Chatbots and Conversational AI: Past, Present and Future

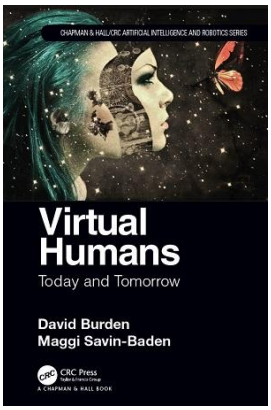
September 2025

DADEN LIMITED
Author: David Burden

INTRODUCTION

This White Paper takes a broad look at the concept of conversational AI. As noted in the sidebar there are a lot of terms that are used for computer programmes which are able to engage us in human-like natural language conversation – one of the most common of which is "chatbot". The topic of conversational AI also has a long history, and the first part of this paper puts the current interest in chatbots into this historic context, describing some of the different approaches that have been taken over the last half-century (!) to develop chatbots. The paper then considers some of the generic benefits of chatbots before looking at different use cases across both business and consumer sectors. The technological foundation of the current generation of chatbots is then considered, looking not only at Large Language Models but also at Knowledge Graph approaches. Some of the broader requirements of a truly "intelligent" and "human" chatbot are then considered, as well as looking at some of the key challenges. Finally, the white paper looks at how the use cases for chatbots might evolve over the coming years, and how they could start to challenge our very sense of what it is to be human.

In some ways this white paper also serves to update part of my 2017 book, Virtual Humans: Today and Tomorrow, as we edge ever close to tomorrow!



Available from Routledge, Amazon and all good booksellers!

What's In a Name?

As the word cloud shows there are a wide variety of different terms applied to conversational AI and the type of systems built with it. "Chatbot" is probably the most generic term, but certainly nowadays has connotations of being a relatively simple system. "Avatar" is also used, but we prefer to keep this to the digital manifestation of the conversational AI. Terms like "Virtual Assistant" or even "Smart Speaker" tend to imply a specific use case.

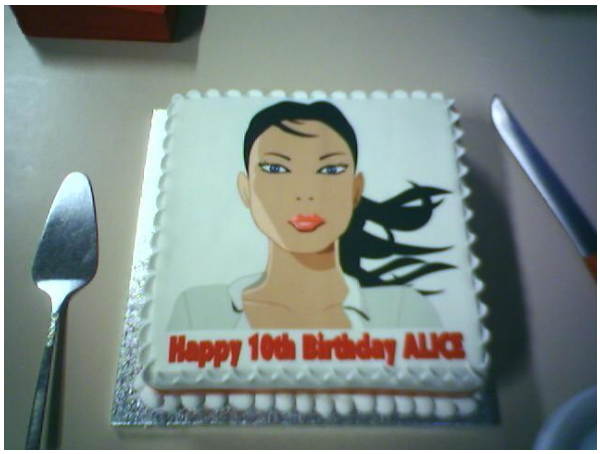
This white paper uses the terms "Conversational AI", "Chatbot" and even just "Bot" synonymously.

A QUICK HISTORY OF CHATBOTS

Early Days and Hobbyists

The development of computer chatbots dates back to at least 1966 when Weizenbaum introduced his **ELIZA** chatbot to the world (Weizenbaum, 1966), (Atwell 2007). **ELIZA** was a "virtual therapist" which had an elliptical questioning style – turning user questions back on the user in an archetypal therapist style.

ALICE, developed in 1995 by Richard Wallace was probably the next major milestone in chatbot development, and was notable for being written in a language – the **Artificial Intelligence Markup Language (AIML)** specifically developed for chatbots and which was understandable by hobbyist coders. **ALICE** used a strict pattern matching approach, although wildcards enables one pattern to match many inputs, and it incorporated a recursive element to enable a multi-phrase input to be split into separate parts for processing, and for a pattern to be rewritten in processing and passed back through the pattern matching system. **AIML** "categories" (essentially a pattern to match and its response) could be shared between different AIML bots, and the core **ALICE** bot had over 50,000 categories (Atwell, 2007). AIML was also implemented on the Pandorabots website, enabling anyone to create their own simple bot – it had been used to create over 325,000 bots by early 2025.



The Alice chatbot's 10th Birthday Cake, produced for an AI seminar in celebration of the event, and attended by Richard Wallace, Hugh Loebner, the author, and many others!

The Turing Test

You can't get far in talking about chatbots without mentioning the Turing Test. Originally called "The Imitation Game" (and based around gender rather than AI), the essence of the test is whether someone engaged in a (typed) conversation on two computer terminals can tell which conversation is with a human and which is with a computer (Turing, 1950).

The test has long been rejected as a test of any form of "intelligence" but is a good behavioural and capability test – just how well can the computer imitate human conversation?

Ironically, one of the biggest mistakes that computers make in the Turing Test is talking too much – human conversation is often knowing about letting other people do the talking (Warwick & Shah, 2017).

Covert Turing Tests #1

One issue with the Turing Test as commonly run is that the judges know they are taking part in a Turing Test, and so may ask questions, or at least look for clues, that they wouldn't do in ordinary human conversation. If a test is staged covertly, so the judge doesn't know that they are taking part, then a far more level playing field is established (Burden et al., 2016).

I've helped develop bots for two covert Turing Tests, and in both cases the bot passed with flying colours – see the following sidebar.

In 1990 Hugh Loebner set up the **Loebner Prize**, an annual competition styled after the **Turing Test** (Turing, 1950) (see sidebar) to find the best standalone (i.e. no Internet connectivity) chatbot (Powers, 1998). The Bronze Prize was awarded for the best chatbot each year, the Silver Prize if a bot ever passed the Turing Test, and a Gold Prize for a bot which could pass the Turing Test in audio and video. The Gold Prize was never awarded, and nor was the Silver, although the Mitsuku bot won 5 of the last 7 competitions and was coming close to the Silver Prize (Worswick, 2018). The last Loebner competition was held in 2019. ¤

In the hobbyist world the next major development was **Chatscript** (https://github.com/ChatScript/ChatScript), created originally in 2011 by Bruce Wilcox to control the NPCs in Sony's **Blue Mars** virtual world. But, when that world was closed Bruce kept **Chatscript** and made it available for public use. Unlike the HTML/XML style of **AIML**, **Chatscript** is a proper scripting language, but so highly optimised for chat and text work that it's very hard to do things like ordinary maths in it. That said it is very powerful, and you can link out to other code in order to connect to other databases and sources. It is hierarchical rules based, with execution dropping down through layers until it finds a rule that triggers, the rule then defining that response. A useful feature of **Chatscript** is concept sets, effectively synonyms, so that for instance ~fruit would match any type of fruit. Many of the Loebner Prize winners in the 2000s were based on **Chatscript**, and Bruce's Rose chatbot was the winner the two times that Mitsuku failed to win in the last 7 Loebner Prize competitions.

Both **AIML** and **Chatscript** were used for commercial and research projects, as well as by hobbyists. Other approaches during this period included:

- The use of **LISP**, a list processing language, with a fairly unique syntax, which was well suited to natural language processing, but really only used in research and academia.
- Scoring based systems, which moved away from the purely deterministic model of **AIML** and **Chatscript** and instead scored the quality of match and then randomly chose from the high-scoring matches. Daden's **Discourse** was one such system (and which twice passed a covert **Turing Test** – see sidebar). ¤

Covert Turing Tests #2

The first test was staged in the virtual world of Second Life and involved "judges" looking at a new virtual shop and talking to the shop assistant. In 50% of visits the shop assistant was being controlled by a chatbot. 78% of participants who talked to the bot, when subsequently surveyed thought the assistant had been human (Gilbert & Forney, 2015).

In the second test, students in a chatroom discussed a topic over three 1 hour sessions. In each of the 10 discussion groups of up to 4 students (and sometimes only 1!) there was a chatbot, pretending to be a student from a partner university. In *none* of the sessions did any of the participants say that there was a bot in the group. In subsequent analysis and interview it was obvious that the bot was an outlier – but still good enough to pass for human (Savin-Baden et al, 2016).

- **VoiceXML** was an XML language developed for telephone call handling and guided the user through a set of dialogs, and could also use a "slot" model to collect information from the user in whatever order they were given (e.g. from, to, date, time) and then access external systems to provide detailed information.

«

Intent Based Bots

Around the late 2010s as pattern-matching neural-net based AIs were beginning to make an impression, most commercial chatbots moved to an **"intent" based model** (Vasquez-Correa et al., 2021). Here the author defined a model question or statement from the user (e.g. "who are you") along with several ways in which it could be asked (e.g. "Who are you", "what is your name", "what's your name") and replied to. Then after building the model the bot would match what the user says with the best match of the various case that it had been given. This removed the need for chatbot authors to define lots of wildcards and synonyms in order to match a wide range of user inputs. How the bot actually responded was then left to the system, and would typically be a list of possible relevant responses, many of which might draw data from a database. Typical systems of this type included Amazon Lex, Dialogflow, IBM Watson Assistant, and Landbot.io. Many of these systems still exist but have increasingly augmented their offering with LLM components.

Many of these also introduced **flowchart style authoring tools** in order to define conversation/business process steps for the bots to follow (Muhammad et al., 2020). As they were typically very task focussed they weren't much used by the hobby/academic community.

Grammatical Parsers

The "holy grail" of chatbot development has been a bot which actually understands human language grammar. These systems tend to be more academic research platforms than hobbyist or commercial systems, but there are libraries in most computer languages to support this approach. One of the most well-known approaches is the **Penn Treebank Part-of-Speech (POS)** tag set, which breaks a sentence down its different elements, which additional code can then use to work out what is being said or asked and how to respond (Taylor et al., 2003). Within such a system the nouns (NN, NNS, NP, NPS) are of particular interest as they tell you

what someone is talking about, and the interrogatives (WDT, WP. WP$) tell the sort of thing that someone wants to know about the noun. POS libraries are in this respect broader than Named Entity Recognition, which typically just identifies proper nouns.

One research finding from discourse analysis is the notion of "**adjacency pairs**" and **"preferred seconds"** – i.e. that people have a particular type of response they expect when asking a question, and it you don't match that response it jars (Sidnell & Stivers, 2012). For instance, if I ask "where is Paris" then I will be more accepting of a response that mentions Paris rather than one which tells me where Berlin is.

The interesting thing is that whilst natural language understanding in this way is challenging, it is (or was) a lot easier to do than trying to do the opposite – natural language generation from raw data.

There has also been some interesting work done about how to assess how well a chatbot talks, with Grice's Conversational Maxims (Quality, Quantity, Relevance, Manner) being one approach that I used a lot (Wilson & Sperber, 2022) (see sidebar). A lot of the newer metrics seem more "technical" or "business" focused, whereas Grice was very user and human-centred.

Grice's Conversational Maxims

Grice's Conversational Maxims are a useful way of evaluating a conversational AI (Wilson & Sperber, 2022). They ask whether a chatbots response:

- Is truthful;
- Provides a helpful quantity of information, not too much or too little;
- Is relevant to the question asked or the topic being discussed;
- Is presented in a clear and orderly manner.

To these I would also add:

- And helps to move the conversation forward.

Even many humans would score low on Grice's Maxims for some of their conversations!

Contemporary Approaches

As of writing (mid 2025), **LLMs** seem to be becoming the dominant approach for building chatbots, although there are lots of cases where hybrid approaches make the most sense, combining the "humanness" of LLM with more managed conversation management (in intent-based systems) or more managed data repositories. Another approach is that of **semantic knowledge graphs (KGs)**. Both KGs and LLMs are discussed in more detail in the section below on **Technology**, and Annex A examines **Knowledge Graphs** in more detail. But for now, let's discuss the benefits of using chatbots.

«

GENERIC BENEFITS

Whilst specific use cases may have their own particular benefits there are a set of generic benefits that almost every deployment of chatbots may attempt to leverage:

- **Reduced Costs** – By using chatbots instead of human staff to manage part or all of a customer or service interaction the overall costs of providing the service, or costs of sale, may be reduced given that once one chatbot has been created scaling it to handle 100s or even 1000s of calls is relatively inexpensive. Human staff could be focussed on the more complex queries, or redeployed to other tasks.
- **24/7 Availability** – Once a chatbot is set up it can operate 24/7, 365 days a year and doesn't need any holidays!
- **100% Consistency** – Once a bot is provided with a script it will stick to it – so every customer or user will get a consistent experience, and the bot will always be "on message" (although that may not be 100% true for LLM-based chatbots with their tendency to make things up). That's not to say that within the boundaries of its programming the bot can't vary and personalize each interaction, or even be creative in how it deals with conversations – but all that flexibility is something that you can define.
- **Mobile First** – Mobile tends to be the access tool of choice for many people for much of the time – but even today's phones suffer from relatively small screens and cumbersome interfaces. Chatbots are actually a more natural fit for mobile than for desktop as they support both voice and text conversations which are "lean forward" modes which people have tended to be less comfortable with in the "lean back" mode of most web applications (Tona, 2015).
- **Anthropomorphic Engagement** – Even though people know that they are talking to a computer, most people can't help but to relate to the chatbot as a person – they even apologise to it if they get something wrong, and thank it when it does what they want. They are also more likely to remember the engagement, and that can help to build the relationship with the organisation and brand (so the experience had better be good!)

Ashley Madison and Tay

Two precautionary tales on the use of chatbots:

Ashley Madison was an online dating agency. In 2015 it was revealed that many of the women that men were chatting to on the site were actually chatbots – designed to keep the men engaged and paying their subscription given the lack of real women on the agency's books (Harrison, 2019)!

Tay was a chatbot designed by Microsoft and launched with much fanfare on the web on 23 Mar 2016 – and shutdown with 24 hours! The bot was a form of echobot, using user input in a pretty direct way to build future outputs. The problem was, internet users being what internet users are, the bot soon started spouting a range of unpleasant views on a range of topics – and no moderation was in place (Mathur et al, 2016; Miller et al, 2017).

- **Anonymous Intimacy** – Users tend to find a conversation with an engaging chatbot, especially one represented by an attractive animated avatar, as more immersive and personal than just filling out a form. At the same time they also feel that they are more anonymous when talking to the bot than when talking to a human being. The net result is something called "anonymous intimacy" (Zhu, 2025), and research has shown that users will often disclose more when talking to a chatbot than when talking to a human or just filling out a form (Savin-Baden et al., 2013).

«

.

HOW CHATBOTS ARE USED

Chatbots are fundamentally a horizontal technology, meaning that they can be applied to almost every type of human and commercial activity. In practical terms they tend to exist on the sort of spectrum shown in

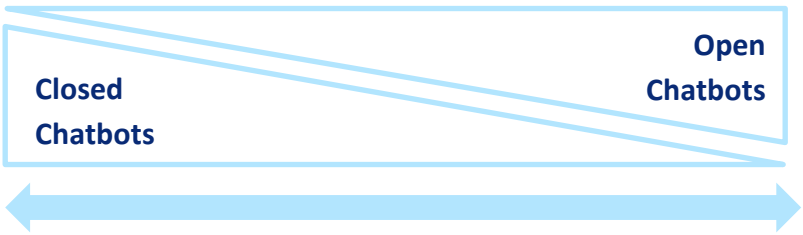


Figure 1, where at one end they are very closed, constrained and task-focussed – only talking about the task for which they've been created, such as booking a flight - whilst at the other end they are very open and can talk about almost anything for as long as you want.

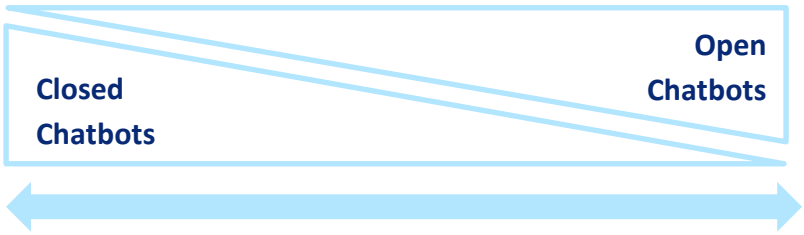


Figure 1: The Chatbot Spectrum

Figure 2 provides a simple landscape map of the variety of industries and roles in which chatbots can be used.

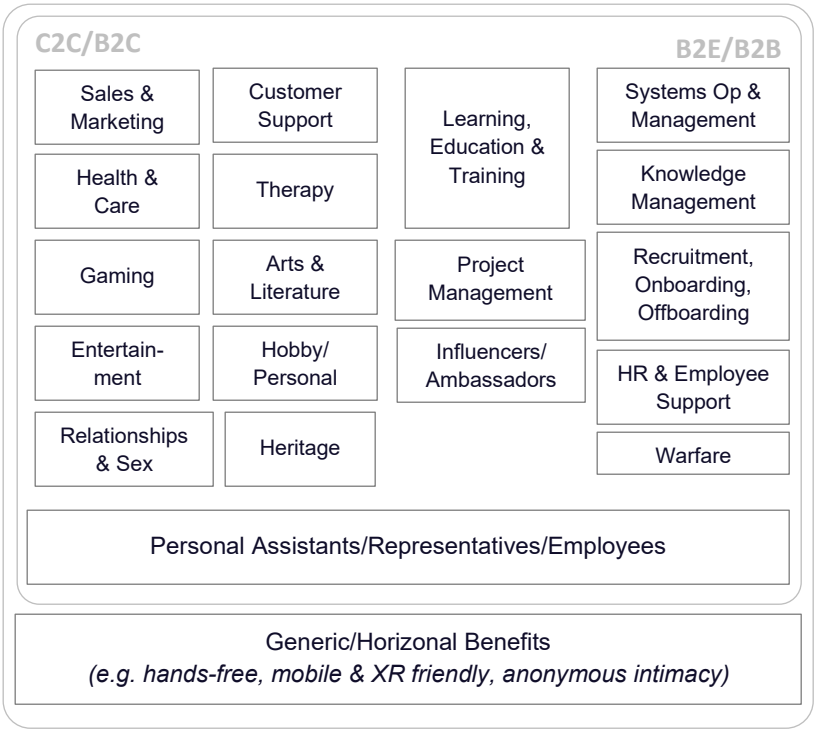


Figure 2: Chatbot Uses Cases (adapted from Burden & Savin-Baden, 2024)

The sections that follow describe specific use cases in these industries and sectors in a bit more detail.

Business to Consumer (B2C) Roles

Retail

The most obvious applications of chatbots are in sales and marketing and customer support roles (Al-Tuama & Nasrawi, 2022; Dangol & Rind, 2024). Typical use cases include:

- Web and phone sales and marketing;
- Virtual world retail assistants;
- Sales calls & follow-up;
- Customer care;
- Supporting Fault-finding;
- Service guides;

There is also an increasing opportunity to use chatbots as service and device interfaces, for instance in cars and home management systems, and of course in smart speakers.

Health and Care

There are significant opportunities for the use of chatbots within health and care (Laymouna et al., 2024; Palanica et al., 2019). These include:

- Remote/virtual diagnosis (NHS111 or even 24/7 GP);
- Remote/virtual management of treatment (e.g. sticking to drug and exercise regimes);
- Remote/Virtual care (especially for the elderly);
- Dementia management;
- Mental health support and virtual counselling;
- Companionship;
- Support for partially sighted and blind people.

None of these are necessarily about replacing trained medical and healthcare professionals, but they are about providing one-to-one, personalized 24/7 support to anyone who needs it, and always with an escalation route to human support.

Games and Entertainment

Since a lot of work on chatbots was done as part of creating **non-player characters (NPCs)** for computer games and virtual worlds (e.g. **Chatscript** for **Blue Mars**), it should be no surprise that there are significant chatbot applications not only in games (as ever more complex and engaging **NPCs**) but also in entertainment.

Whilst currently more about scripted performances rather than two-way dialogue the current use of talking avatars in video production (e.g. HeyGen), podcasts (e.g. NotebookLM) and as virtual DJs (e.x. "X" on Spotify) could naturally evolve into two-way chatbots – either to deal with questions and build engagement after a production, or to increasingly make the production itself more of an interactive engagement.

This could also extend into more "high-brow" entertainment. LLM-RAG based systems are already able to turn a dry text book (or even a novel) into something you can talk to, and projects such as Echoborg (Eagle et al., 2021) and eDrama (L. Zhang et al., 2009) have examined how chatbots can be used in a dramatic role.

Hobby Bots

Ever since the days of ALICE and the Pandorabots website, hobbyists have been keen to create their own chatbots – often just for the joy of playing around with the technology, but equally often to explore historical characters, create virtual copies of themselves or media personalities, act as virtual sales and support staff on small websites, or to act as personal brand ambassadors, science fiction characters or virtual friends. Pandorabots is still around (https://home.pandorabots.com/home.html), and still using AIML. A more modern offering is Kolz (https://kolz.chat/), built on the Sensay (www.sensay.io) LLM-enabled platform.

Heritage

We are all no doubt familiar with the recorded audio tours you get in museums, galleries and heritage sites. Chatbots can not only deliver this passive tour, but also be ready to answer any questions that the visitor may have, and even to take on the persona of characters in a scene or painting, or of the artist themselves, in order to provide the visitor with a more engaging, immersive and informative experience (Nafis et al., 2022).

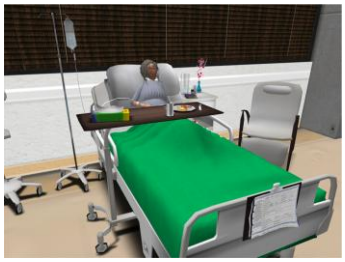
Tourism and Real Estate Guides

Both the tourism and real estate industry have an interest in having guides show you around places – be it a new city, a heritage site (as described above) or a potential new home. A chatbot can provide that service 24/7, implemented either as a text , audio or video guide on your phone, or as an avatar within a 3D virtual or mixed reality experience (Benaddi et al., 2024) (Szumilo & Wiegelmann, 2024).

Sex & Relationships

Even a cursory look at the chatbots on hobby/fan sites such as Pandorabots will show how many bots are created as the "perfect girlfriend" (perfect boyfriends seem a lot rarer – but see Japanese *otome* games)(V. N. Zhang & Vlachokyriakos, 2025). As GAN style approaches to video production improve, not to mention the ability to deploy chatbots as full body avatars into virtual worlds, then using chatbots to deliver the full girlfriend experience (GFE) becomes an increasingly lucrative proposition – *Blade Runner 2049* pointed the way to this with Joi – Officer K's virtual girlfriend (Villeneuve, 2017) (Nugraha & Noorman, 2021). In fact, such virtual relationship (or even sexual) partners have a long history – with virtual worlds such as

Chatbots as Educational NPCs



A lot of the education and training projects I've been involved with in virtual worlds have used chatbots to create in-world characters for the students to interact with. These include:

- Mable, a dying patient for students to practice end of life care procedures;
- Citizens of New York taking refuge in a hurricane shelter – and presenting an array of issues for managers to deal with;
- Various nursing simulations, including dealing with a Post-Partum Haemorrhage, a drugs round and a patient transfer
- Client and counsellor for students to observe in counselling training.

Red Light Centre offering virtual sex avatars back in the 2000s, and, as discussed in the earlier sidebar, the Ashley-Madison website used chatbots as virtual female clients to help recruit more male clients. Chatbots, especially linked to animated video or virtual world avatars, could even allow OnlyFans creators (and other sex workers) to extend their offering (Gentleman, 2025).

There may even be a role for chatbots in virtual dating. Rather than just create a profile on a dating web site why not upload a chatbot replica of yourself that potential dates can talk to – and which can of course report back to you. It may not be long until that first contact is actually between your chatbot and their chatbot!

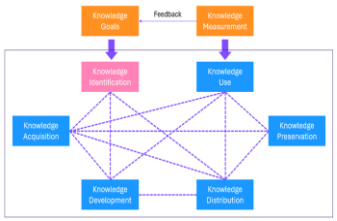
Needless to say there are significant ethical and moral issues with a lot of these use cases, particularly around the objectification and exploitation of women (Borau, 2025; Davey, 2023; Depounti et al., 2023; Koh, 2023; Wu & Wang, 2025). There is also the very vital issue of chatbots (and particularly video enabled chatbots) by unscrupulous and/or malicious people or organisations whether it's for commercial porn or romance fraud (Veisi et al., 2025). Less extreme cases though are still of concern, particularly when behaviours such as anonymous intimacy are considered, and these are discussed in more detail at the end of this white paper. «

Training & Education

Training and Education is considered separately as it includes elements of both B2C and B2B/B2E. There have been numerous projects looking at the use of chatbots in an education and training context (Hwang & Chang, 2023; Kuhail et al., 2023; Nee et al., 2023; Pérez et al., 2020). Key use cases include:

- **Virtual Tutors – topic focused**, so essentially like a virtual subject expert
- **Virtual Tutors – leaner centered**, so more like a tutor in the historic sense, getting to know the learner really well, and their learning preferences, and then directing them to content and encouraging and support them through it accordingly.
- **Virtual Trainers** – more based on skills delivery, particularly when deployed either in VR with 3D models of any relevant equipment or environments, or deployed in MR/AR in front of the user, again with supporting 3D models or overlays on the real equipment.
- **Virtual Mentors** – more about longer terms and pastoral care, ensuring that students stay motivated, plan their learning, dealing with problems, and putting their learning

Chatbots and Knowledge Management



Within a business and organisational context it can be useful to think of the role of chatbots as just another part of the organisations knowledge management system (Probst & Romhardt, 1997). In particular, conversational AI can have a significant a role in knowledge:

- Acquisition;
- Distribution; and
- Preservation.

It is also increasingly having a role in Knowledge Use (welcome back Mr Clippy!), and in future will no doubt play a role in Knowledge Identification and Development.

into a broader context. There may be some overlap here with mental health support.

- **Non Player Characters (NPCs)** within training environments, acting as for instance virtual staff, customers, team members or patients (see sidebar). One particularly interesting piece of research found that paramedics learnt and performed better in virtual training when there were virtual bystanders standing around watching – just as there were likely to be in real life (Rovira i Pérez, 2016).« (Probst & Romhardt, 1997)

Business to Business (B2B)

Many of the sales and marketing roles within B2C can also be found in B2B, and chatbots can be used in all the roles described there – although their use is likely to be more common at the commodity end of the market, with high value sales still having an emphasis on human relationships (but increasingly AI supported). Support and maintenance roles are also just as viable as for B2C, and high value products probably justify the investment in very capable, 24/7, location independent chatbot support.«

Business to Employee (B2E)

There are significant opportunities for chatbots with Business to Employee (B2E) services. These include:

- **Recruitment** (including briefing/self-selection, retention and motivation during recruitment, interviewing and role familiarization/day-in-the life briefings) (Barghi et al., 2022; Koivunen et al., 2022);
- **On-boarding** and personalised support for the first few months (Kylliäinen, 2024);
- **Training & Education** (see above);
- **Employee Support** (pastoral care, mentoring) (Roberts et al., 2020);
- **Corporate knowledge management**, providing conversational access to ideally the sum total of the knowledge in the organisation (Sherif et al., 2024). As discussed below there may even be roles for project and product specific chatbots as the foci for knowledge (and even co-ordination) around specific areas.

«

Platform Personas

Whilst Rommie from *Andromeda* might embody the worst of the female avatar/chatbot tropes, the concept behind her/it is a very sound one. What if your complicated system or platform – be it a factory or warship – was represented or fronted by a conversational AI – one which could be present on Zoom calls or in virtual or mixed reality sessions. You wouldn't ask a colleague how the system was – you'd ask the system itself. And the system/AI would build up years and years of experience in how it works, what can go wrong, how to fix it, and how to get it to operate most efficiently and effectively.

Industry 4.0, 5.0 and the Industrial Metaverse

Industry 4.0 is all about smart factories, where digital and physical systems are interconnected, automating and optimizing manufacturing processes (Golovianko et al., 2023). Conversational AIs can help navigate the complexity of such systems, enable natural dialogues to understand the health of systems and flow, ask for predictions and recommendations. They can play even more of a role in **Industry 5.0** (Golovianko et al, op. cit), where the core tenets are an emphasis on personalisation and sustainability, with a shift in focus to human–machine collaboration. This effectively becomes an instance of the "platform as persona" model where the factory or industrial process itself becomes a persona with whom its human colleagues can interact with as though talking to another person. Pushing on, the **Industrial Metaverse** takes this further by creating virtual environments within which not only can the systems and flows or digital twins of an industrial process be visualised and examined but the factory persona can be present as an avatar themselves, showing people their systems, explaining what is going on inside them and effectively providing an industrial equivalent of the Rommie ship avatar from Gene Roddenberry's *Andromeda* (*Andromeda*, 2000) (see sidebar). «

Science and Research

One of the classic models of a conversational AI is the Librarian character in Neal Stephenson's seminal *Snow Crash* – which gave us the Metaverse (Stephenson, 1992). The Librarian is the spiritual forerunner of ChatGPT, an avatar in a virtual world which can answer almost any question, and help find the answers to those it does not know. Tools like NotebookLM are already providing researchers with the ability to build powerful conversational AIs from a variety of research sources. Now imagine that every research lab had its own persistent AI built from every bit of research ever done at the lab, or every journal had its own AI that knew about everything ever published in the journal, or that agentic AIs could be given research tasks in the way that they might currently be handed out to a research assistant. Of course there is the challenge here, as with other areas, that if the AIs pick up the "assistant" roles then how do we ever train the next generation of "professional" researchers.

Paintings or Daemons?

In talking about virtual assistants and virtual personas one model I've always found useful is that of Dumbeldore's paintings and Philip Pullman's daemons.

At Hogwarts, if Dumbledore has a problem he can chat to the paintings of previous headmasters and headmistresses. There is a wealth of experience and wisdom there – but each has its own biases, gaps, and beliefs.

In contrast, the daemon's of *His Dark Materials* represent considered best practice, a conscience on your shoulder, advising on the best way to do things.

Warfare

My white paper *The Military Applications of Conversational AI* (D. J. H. Burden, 2025) goes into a lot of detail on potential military uses of conversational AI, many of which have their parallels in industrial and business uses. Some of the roles include:

- Training & Mentoring – as trainers, tutors, and NPCs;
- Virtual Staff Officers;
- Commanders' representatives;
- Platform personas; and
- Project personas.

My book *The Military Metaverse* (Fawkes & Burden, 2025) even considers what a war in an AI avatar populated metaverse might be like.

Personal Assistants

The main cross-cutting application of the chatbot, and one which is only slowly becoming actually useful is that of the **virtual personal assistant (VPA)**. Smart speakers are a relatively passive starting point for the technology, and systems like Siri have never quite taken off in the way that some expected. But the latest generation of LLM chatbots, for instance ChatGPT 4o running in hands-free mode are getting a lot closer to the VPA dream, particularly if they start to get access to your email, IM, video conferencing, diary and other systems. Another route towards the VPA could come from broadening the development of **Virtual Life Coach/Virtual Mentor** systems, which start with the soft-skills, pastoral support, and then augment it with more task orientated capabilities.

Some more advanced applications of chatbots will be considered later in this white paper.«

TECHNOLOGIES

In this section I'll look in a bit more detail at the range of technologies that can be used to implement and support conversational AIs.

A useful diagram, showing the components elements of a virtual human, reproduced from my book *Virtual Humans* (Burden, 2019), is shown in Figure 3.

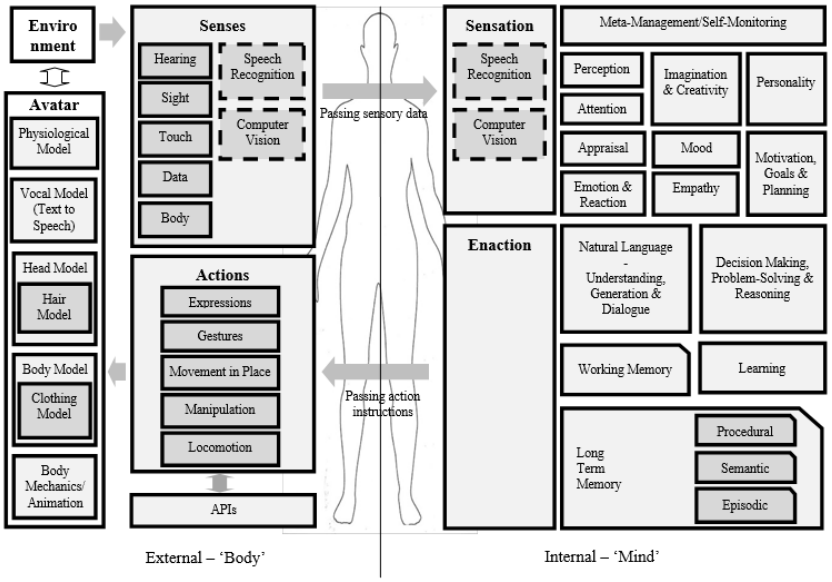


Figure 3: Components of a Virtual Human (after Burden & Savin-Baden, 2019)

The left-hand half shows the elements needed to represent a virtual human (or conversational AI) in a digital environment, and the right-hand side shows the systems that go into making the virtual human's "mind". Generally, the representational elements are a lot easier than the mind elements, and increasingly we can probably consider that a conversational AI can be represented in whatever form we need – a simple text chat interface, an audio or audio/video interface, or an avatar within a virtual reality world or projected into our physical reality through a mixed reality system.

Far more challenging is what goes into the mind. A traditional chatbot acting as a glorified question answering system really needs nothing more than a natural language capability. However, the experience of talking to an LLM chatbot can be so lifelike that we expect it to show many more of the elements of mind than a "simple" LLM enables. If conversational AIs are going to deliver on expectations, and certainly to excel in all of the roles described earlier, and those more advanced roles described later, then we are

likely to see them evolve more and more towards being true virtual humans, or at least virtual beings.

For a full description of each of the components pick up a copy of *Virtual Humans*. In this white paper we will focus initially on the chat function, and then look at how some of the other elements of mind could be introduced.

«

Substrate Independency

One principal of consciousness and sentience that many (including myself) believe in, but many others don't, is that of substrate independence. That is that there is nothing special about the brain that enables consciousness or sentience – it's all a matter of processing. There are no "spooky" effects from quantum mechanics or panpsychism or Integrated Information Theory fields. As far as I'm concerned consciousness is ultimately reducible to code (albeit possibly very complex, and probably very elegant code), and once you understand that you can run it on anything.

Chat

At the start of this white paper we looked at how chatbot systems have evolved over the last 50 or so years. One interesting aspect is that until recently the focus was on **Natural Language Understanding** (NLU) systems, but generating human-like text from a set of data generated for the response (**Natural Language Generation**) received far less attention. With LLMs significant strides seem to have made in both areas. However, as intimated earlier, LLMs are not without their problems, and whilst they may well be a part of any conversational AI system going forward, particularly with approaches such as RAG (see below) to constrain knowledge to particular domains and knowledge sets, they may not be all there is.

LLMs

OpenAI was founded in 2015, and in June 2018 released **GPT-1**, with the first conversational interface, **ChatGPT** being released in November 2022 (Marr, 2023). LLMs (including **ChatGPT** and competitors such as **Claude** and Meta's open-source **Llama**) work by analysing huge quantities of text and then building a model of, essentially, what words follow what other words. However, building this model takes significant amounts of processing power and time. Once done, the LLM is able to generate reasonable responses to most questions, can generate new creative content, and go well beyond being a simple auto-complete, parroting or question-answer machine. A "pure" LLM system cannot though tell you where it got the information from for its answer, the answer just "emerged" from its database of word vectors, and is also liable to "hallucinate", i.e. make answers up – although a lot of work is going into reducing or eliminating hallucinations, and also to involve humans in the generation of its corpus and fine tuning of code and responses in order to help create better conversations. Hybrid systems using LLMs are also increasingly common. For instance, Microsoft's **Co-Pilot** essentially deals with questions by doing a Bing search and then using an LLM to rewrite the output as a text summary –

The TAO of Topic Maps

A good introduction to the idea of topic maps is the TAO of Topic Maps (Pepper, 2000). Here TAO represents how topic maps work, they have:

Topics – such as "cars"

Associations – relationships between topics, which make up the links in the mind map, e.g. Cars made of metal

Occurrences - external information about a topic, e.g. a picture, video, report, URL etc

Note that "occurrences" can sometimes be used to means "instance" – so GD69SGD (a car reg) is an instance of the topic "car". Topic maps can also have "facets" – essentially meta-data (often numeric) about a topic – for instance mass or date of creation.

complete with the links to the source data, something that has now become common at the top of all Google searches. Another common extension is that of **Retrieval Augmented Generation (RAG)** (Church et al., 2024), where an organisation (or individual) can supply their own documents to be "indexed", and answers are generated initially by using the information in those documents, again with the LLM rewriting the output into usable text. The most recent (mid 2025) move has been into **Agentic systems** (Miehling et al., 2025), where LLMs bots are able to chain a whole bunch of steps, making use of a variety of different API connected systems in order to complete a task.

«

Knowledge Graphs

One of the most interesting alternative approaches to chatbot development, and to knowledge management more generally, is that of **knowledge (or semantic) graphs** (see Annex A)**,** which grew out of ideas such as **mind maps** and **topic maps** (Pepper, 2000) (see sidebar). Here the "atomic" unit of knowledge is known as a triple – built of a subject, predicate (verb) and object, so:

```
Car  madeof     Metal
```

These triples are then linked together in a knowledge graph, so that, for instance, the "car" node links to all the things known about cars (including their uses) and to all instances of a "car", the "metal" node links to everything known about (and made of) metal, and each instance of a metal, and both inherit information from the fact that a car is a form of transport and a metal a type of substance, and can both donate information to child objects (such as instances), so that once you know that GD69SGD represents a car, or that iron is a metal you can infer a lot about them.«

This idea about inheritance is very powerful, as if a user tells a bot that they were involved in project X, the bot immediately knows to ask them all the things it typically knows about projects (which can be defined in a "prototype" project), such as when it started, when it finished and what the objectives of the project were. Using this approach we can (and have) completely bootstrapped a bot from a few simple questions such as "what projects have you worked on", and then use the prototypes to collect all the information, which inevitably mentions more projects, people, organisations etc that the bot could then ask more questions about. Figure 4 shows the core

prototype graph that we used on the **Virtual Barry** project (D. J. H. Burden, 2019), and

Figure 5 shows just part of the knowledge graph built for a real-world project.

«

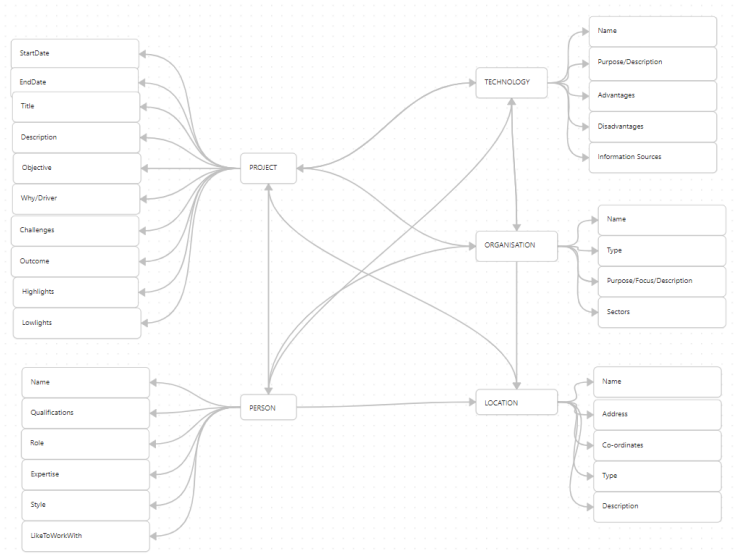


Figure 4: A Skeleton Knowledge Graph for someone involved in Tech Projects

Of course, if building an ontology properly every item should be able to trace a parentage back up to some global "thing". There has been a lot of work done in creating so-called "**Top Level Ontologies**" (Jansen, 2008) (see sidebar) to provide a global context (and common sense) for any underlying ontologies. Although not strictly a knowledge graph or ontology, **WordNet** is probably the best known example of the form (G. A. Miller, 1995). In this new world of LLMs such TLOs may still have a role, and may help in bots better understanding context and in being able to better situate new information. There may also be a role for LLMs in creating improved TLOs. «

Top Level Ontologies

A Top Level Ontology is a bit like a huge game of animal-vegetable-mineral. You start with "thing" and then split it down into the main subcategories (probably "physical" and "non-physical"). You then split these down (e.g. physical into organic, inorganic and mixed), and so on. Given that levels are essentially going up in powers, if you have, say 5 things at each level, then in just 9 levels you'd have nearly 2 million categories. At every level you define all the things that are unique about this category, and what ought to be known about things in it (the prototype). Every "instance" of a thing in the real world would then slot into one (or possible more) of these categories, and through inheritance you should instantly know a lot about it, what relevant questions to ask to know more, and how it relates to other things.

Using Knowledge Graphs

One of the most powerful uses we found of the graph was having the bot respond with the "best match" triple for a query, but if the user just said "more" (or nodded their head) then the bot returns an adjacent triple on the same topic, and then with repeated nods might switch the current topic to one connected to the first one, and so just gradually crawl over the map. Setting the parameters for speed of "crawling" was a good way to simulate different personality types!

The graph was also excellent for finding paths (i.e. linkages) between entities which even the original person didn't think of and identifying clusters of related topics. In fact, it's the "mining" of this graph that begins to open the way to the bot creating original insights from the data in a way probably not possible with LLMs.

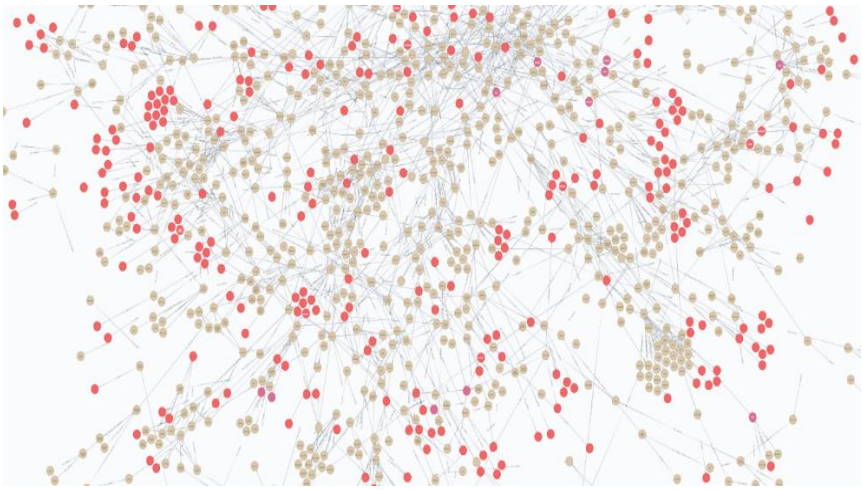


Figure 5: Just part of a real-world knowledge graph!

One of the best books we've found on the whole topic is *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL* (Allemang & Hendler, 2011). «

Why Knowledge Graphs?

One of the key problems with LLM-based systems is that they don't actually "understand" anything and so have limited ability to reason. It is also hard to overlay meta-information on the information that they work with and generate. With a knowledge graph approach the graph *is* the understanding of the information. As such the AI can reason about the information, we can easily add meta-data to the information (even to the links), it can find patterns and clusters and outliers, it can find connections which may not be obvious to humans, it can potentially fill gaps in the data and as described above it can use the links to guide conversations and enable more human-like conversations. But building knowledge graphs can take a lot of effort, and turning **SPARQL** queries (see Annex A) into interesting and varied text can be non-trivial, but luckily these are both areas where LLMs can be of assistance. Long-form text site Medium, in particular, has some good commentary on the intersection of LLMs and Knowledge Graphs (*Artificial Intelligence in Plain English*, n.d.; *Neo4j Developer Blog*, n.d.).

Hybrid Systems

All of the conversational AI approaches described above have some merit in modern systems. Intent-based systems can very easily deal with more basic queries, LLMs are very effective at both understanding and generation, but are susceptible to hallucinations

and hard to control, knowledge graphs have the structure and rigor to support understanding and reasoning, but need something to front them, and grammatical parsers can help when all else fails in trying to understand what the user is trying to say, and give some clues as to whether a question is best dealt with by a intent, LLM or knowledge graph type system.«

In my opinion, given the current state of technology, the best approach to building a powerful conversational AI is a hybrid one, where a core "control" application "husbands" the conversation, working out which back end systems are best placed to respond to a given input, orchestrating the use the use of multiple systems (sequentially or in parallel) to complete a task, and to provide some level of conversation and even multi-session management (building on areas such as discourse analysis and adjacency pairs) to lift the bot from a glorified one question – one answer system to something which can have a meaningful and directed conversation over multiple sessions, with clear intent on both sides. This idea of husbanding aligns quite closely to some of the current ideas around how agentic systems should work (Finio & Downie, 2025).

Socially Aware Bots

Chatbots are usually thought of in terms of one-on-one conversations, but they will increasingly be used in multi-user environments, whether that's a chat room, a voice or video conference call, or a social virtual world or mixed reality space. In these situations, the chatbot needs to have some additional technical, and social skills. For instance, it needs to understand turn-taking, to know when it's appropriate to speak, to know when it's being addressed, and if in a crowded virtual world environment be to able to separate out the different clusters of people talking and only respond to its group. Research has already been undertaken in all of these areas, but more is needed.

Memory

A key part of any mature conversational AI is going to be its memory. Raw LLM systems currently have very limited ability to remember things, and are typically augmented by some form of external storage service. Given the hybrid approach proposed above, I would argue that most of the bot's memory for new information should be in the form of a knowledge graph.

Cognitive scientists usually divide human memory into **working memory** and **long-term memory**. Working memory is that handful of facts you remember whilst working through a conversation or task, long-term memory is everything else. Working memory in a chatbot is typically managed by the core system code, such as remembering the user's name, the current topic being discussed and so on.

Scientists divide long-term memory into 3 main elements, and it makes sense for a chatbot to reflect these in how it manages its memory. The elements are:

- **Semantic Memory** – our knowledge of "things" and how they relate to each other – the essence of what a knowledge graph represents.
- **Episodic Memory** – the memory of our own experiences, what has happened to us, and the emotions and anecdotes related to them. All of this can be readily coded into a

knowledge graph and linked to semantic memory so that whenever a semantic node gets discussed any related episodic memories are there for use.

- **Procedural Memory** – our memory and knowledge of how to do things. This tends to be more linear and stepwise than other forms of memory, and whilst it can be shoe-horned into a knowledge graph may be better managed through other structures (e,g, XML or JSON), but any entities can still be linked to relevant parts of the semantic and episodic memory knowledge graphs.

One of the challenges with memory (especially episodic memory) is knowing what to store – as if you store everything (what happens to you every second) then volumes can become excessive. One approach for a chatbot is to mimic some of the human processes, for instance:

- Storing memories with an emotional valency attached, so more emotional memories are stored for longer (or are easier to access) than less emotional ones;
- Having generic memory templates for things that are very routine, so the detail of every instance isn't stored, just the variations from it;
- Memories can be given a "half-life", and become harder to recall the older they are, modified by emotional valency, till they just get "forgotten";
- Using "sleep" as a time to sort memories and even to generate "dreams" for the AI whilst it does so (see below).

«

Representation and Sensation

As mentioned earlier, the working assumption in this white paper is that a chatbot or conversational AI can be presented in whatever medium is required. This includes:

- **Text chat** – where as well as language a bot should also be able to use and understand emoticons and of course to understand the meaning of TALKING IN CAPITALS!
- **Voice** – where the bot should be able to read the tone and emotion of the user, and to set its own tone to reflect its emotional state (see below), and to show empathy.

Virtual Embodiment

One aspect of bots which I think will prove to be absolutely essential for building anything approaching a sentient AI is that they need to be embodied. Being embodied provides grounding in knowledge – it no longer becomes theoretical, and notions such as embodied cognition are explored in *Virtual Humans* (op. cit.).

However, this embodiment doesn't have to be physical – and in fact by making physical robots and androids we have to deal with all sorts of electromechanical and physics issues which are secondary to the key one of embodiment. For bots the best way to embody them is in virtual worlds, as avatars.

One big advantage of this approach is that it provides a level playing field between humans and bots, both are represented as avatars, and both can have the same agency in the virtual world (Burden et al, 2016).

- **Video** – as well as voice the bot should be able to read the expressions and gestures of the user, and to be able to control its own expressions and gestures. Any video avatar should avoid the "uncanny valley" where the animations just look creepy, better to stay cartoony or implement a really good animation system (which with systems such as HeyGen is becoming a lot easier).
- **Avatar** – in virtual reality and mixed reality the chatbot can be represented as a full-body avatar. Again, uncanny valley should be avoided and the chatbot should be able to read and display a full range of expressions, gestures and movements. One of the big advantages of virtual worlds has always been that everyone is an avatar so the chatbot has a level playing field when it comes to trying to be just like any other (human) user (D. J. Burden et al., 2016)
- **Robot** – where a physical form is required (see sidebar) (Egan, 2010) «

As soon as the chatbot is represented in an environment it should ideally be able to sense that environment as well as any other user. In text chat that's just reading the chat (and case, font, even pauses). In audio it might need to be able to identify any extraneous noises ("are you at a party", "are you in a car", "are those birds?"). With a video call it may need a full range of image and video recognition – particularly if someone starts showing slides or pictures, or holds up an object. In a virtual world the bot should be able to sense and comprehend the entire environment, and everything and everyone in it. If manifest in the real world as a robot then it may also need to be able to detect smells, understand touch, and perhaps even taste!« .

Androids and Gleisner Bots

The classic science-fiction trope of an intelligent android walking around with its "brain" in its head it is probably an unrealistic one. The android body is just too vulnerable, and too limiting, for it to be the sole existence of an AI, such as a well developed Replica. A more likely scenario is one where the AI exists "in the cloud", probably in a distributed and multiply redundant (even fractal) from, and then downloads the relevant parts of itself to one (or more) android (or other body, or even craft) as required, and which are capable of operating autonomously. It then dynamically or periodically synchronises the memories and capability enhancements across all of its manifestations. Such a model is shown in the Gleisner robots in Greg Egan's Diaspora (Egan, 2010).

EXTENDING THE MIND

Although some of what has already been discussed may seem like science-fiction, we are really only at the start of what conversational AIs may be, and in some cases already are, like. In this section I consider the capabilities which can be added to relatively simple chatbots and conversational AIs to make them even more powerful and human like, and which could ultimately put us on the road towards sentient machines.

Beyond Recall - Opinion/Advice/Deciding

Most chatbots are predicted around simple information/fact retrieval and recall. That is slowly changing with LLMs, but it's worth thinking about the different types of enquiry that we might want to make of a bot (as of a real person), so that we can ensure that the bot is best able to respond to them. Some key enquiry types are:

- **Opinion** – what does the bot think about something, often around likes or dislikes if representing a person;
- **Advice** – helping someone think about something;
- **Suggestion** – helping someone make a decision, e.g. what someone could do about something;
- **Recommendation** – recommending a particular course of action, e.g. what someone should do;
- **Confirmation** – seeking the bots agreement that a chosen course of action is the right one;
- **Prediction** – what does the bot think will happen in a given situation.

Key features of these are summarised in Table 1. Each of these can also be considered in terms of how the bot operates when it does or doesn't have good prior knowledge of the topic, and when asked about things which post-date its own knowledge base. Some of the these will be considered further below when talking about decision-making and wisdom.

Human Machine Teaming and Shared Mental Models

Whilst human-machine teaming (HMT) is much discussed, a lot of it is in the context of the "machine" being a relatively dumb tool. The assumption here, though, is that conversational AIs are possible near-equal partners of human team-mates.

Studies of elite teams show that their high performance is often due to them having the same shared mental model (SMM) of the task at hand (Richards et al., 2018). We all have our own mental model as to how things, or tasks, work. The trick for a human, or human-AI team to perform well is for all the participants to share the same mental model of what needs to be done, how, and what each individual's responsibilities are. Work on integrating AIs and SMMs is already underway by Daden and others (Lyons et al, 2021).

Model	Purpose	Nature	Basis	Tone	Actionable	Output
Opinion	Sharing a personal view. What I think.	V. Subjective, Dichotomies (good/bad)	Personal belief or feeling	Subjective/ informal	No	No decision/influence needed
Advice	Helping someone make a decision. What you might do.	More about ideation. Do/Don't.	Experience or knowledge	Informative, supportive	Sometimes	Merits, assistance, but not recommendation
Suggestion	Helping someone make a decision. What you might do.	Ideation with selection	Experience or knowledge	Informative, supportive	Sometimes	Single/small number of ideas. Little endorsement
Recommendation	Promote a course of action. What you should do.	Expect more objective, a model/rationale/heuristic for the decision	Analysis, expertise or evidence	Direct, formal, persuasive	Yes	A specific course of action
Confirmation	Endorse a specific action/view	Take existing idea/course of action. Fairly subjective?	Personal belief or feeling, or experience	Subjective/ informal	Usually	Validation of idea/course of action
Prediction	Predict the result of future action. What may happen.	Projecting into the future. Ideally needs some sort of model else highly subjective.	Expertise, experience or feeling/personal belief	Informative	No	Prediction, future change in states

Table 1: Beyond Recall - Different Chatbot Enquiry Modes

Perception/Attention/Appraisal

Although not a big issue when chatting to a bot one-to-one through a text-chat interface, perception, attention and appraisal are vital when the bot in embedded as an avatar in a virtual world, or is interacting with the world through a video feed. What are the elements of the scene that the bot needs to perceive, which of those does it need to pay attention to, and how does it evaluate the importance of what it sees. For instance, in a video call should the bot only perceive or attend to the person it is talking to, or should it be aware of the rest of the scene behind the person. The situation becomes more acute when the bot is able to hear (perceive) multiple conversations, again should it filter out all but the main one it's

engaged with or should it maintain a human-like “cocktail party” effect and raise its awareness only if another conversation mentions its name or a topic of interest?

Agency

In order for a bot to be truly useful it needs some level of agency. In current terms this may mean that it has “agentic” capabilities – i.e. the ability to pass instructions to other systems in order to gain additional information, process some data, make a booking, even buy something. In virtual worlds a bot avatar should ideally have as much agency as a human avatar, being able to move things, build things and even script things in the space. As an example, in an industrial metaverse you should be able to discuss a process with the factory persona and then have the persona make any changes or orders that you discuss.

Of course, you may want to categorise levels of agency such that the bot has authorization to do some things without human intervention, but needs human authorization for others. A good model for this is **Sheridan and Verplank’s 9 levels of system autonomy** (Sheridan & Verplank, 1978), (Parasuraman et al., 2000).

Automation Level	Automation Description
1	The computer offers no assistance: human must make all decisions & actions
2	The computer offers a complete set of decision/action alternatives, or
3	narrows the selection down to a few, or
4	suggests one alternative, and
5	executes that suggestion if the human approves, or
6	allows the human a restricted time to veto before automatic execution, or
7	executes automatically, then necessarily informs humans, and
8	informs the human only if asked, or
9	informs the human only if it, the computer, decides to.

Table 2: Sheridan & Verplank's Nine-level Autonomy Model (Sheridan & Verplank, 1978)

Emotion and Reaction

Linked to perception, attention and appraisal is how the bot shows emotion and reacts to situation. In prior work I implemented the **Emotional Alert Detection Systems** architecture developed at the University of Wolverhampton into a bot in Second Life (Slater &

Burden, 2009). As well as hard coding whether some objects caused happiness we also configured the bot so that if objects (e.g. a snake) damaged the bot's avatar through Second Life's in-built damage system then it learnt that seeing a snake could be bad and would show fear. We also implemented a high-road/low-road reaction system, the low-road causing the bot to throw up its hands in horror and step back, the high-road being routed via its conversational AI and causing it to say something along the lines of "look out, there's a snake". Many researchers, myself included, use **Ekman's standard 6 emotion model** (anger, surprise, disgust, enjoyment, fear, and sadness) to create a basic emotion engine (Ekman et al., 1969). It can also be useful to think about these on two sets of axis, one about self vs others and the other about present/past vs future. As mentioned above, when laying down memories it can be useful to store the dominate emotion that was present, and this can help determine how quickly a memory fades – and even if it fades at all.

Mood

In academic research "emotion" often only refers to the instantaneous reaction to something. What we more colloquially talk about as longer term emotion is instead call mood. As well as moods which reflect those basic emotions there are other elements of mood which reflect some aspects of personality – such as being in an introverted/shy or extroverted/party mood. A sudden change in emotion can act as an impulse into mood, causing a mood to change in its level (valence), and perhaps to new mood to dominate for a while, but then over time the mood reverts to the person (or bots) quiescent state. I've found that giving moods half-lives can be a useful approach, so that the mood's offset from its quiescent value halves every minute/hour/day/week or whatever, depending on the strength of the trigger, or the bot's assumed emotional stability.

Another useful approach I found to giving a bot the illusion of changing moods was to give it a set of **biorhythms**, metrics which change over different time periods and reflect physical (23 days), emotional (28 days), and intellectual (33 days) energy.

Empathy

Creating a bot with some form of emotion model is relatively simple, but trying to create a bot which shows empathy is far more complex. The bot needs to be able to detect emotion, and then adjust what it is saying and how it is saying it accordingly. With LLMs this second half can probably be accomplished by a simple pre-

prompt that says something along the lines of “as though you are talking to someone who is feeling sad.....”, but the bot needs to detect the sadness in the first place. There are a number of possible ways of achieving this which show promise, from expression and mood detection from facial and voice analysis, to detecting pupil pulse changes (from video or dedicated sensors), or galvanic skin response. Even the choice of words may give a clue to someone’s mood. In prior work I’ve used galvanic skin response to detect the level of stress that someone is feeling and have a tutor bot adjust its tuition accordingly.

Personality

Personality probably sits at the foundation of an emotion-mood-personality hierarchy, and defines the longer-term form of a person’s actions and interactions. Personality isn’t static, but probably changes over years, if not decades. Again, there are some useful, if “pop” psychology models that can be used to help define a bots personality – the most obvious being the “**Big 5**” model (Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism) (Soto & Jackson, 2013) and **Myers-Briggs** (Pittenger, 2005). In many cases we want the bot to show a minimal, but usually helpful, agreeable and polite, personality and not vary from this. The personality model can also set the bots sensitivity to changes in mood, which can in turn set its sensitivity to emotions.

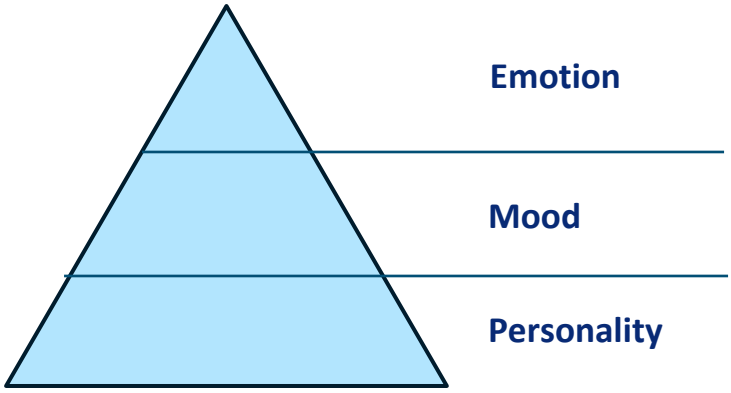


Figure 6: The Emotion-Mood-Personality Relationship

Intent, Motivation, Goals and Planning

If a bot doesn’t have any way of working out what it wants to do – whether at the level of a conversation or of its whole life (whatever that is) then it is condemned to being a reactive agent. Whilst this

might be what is needed in many roles (do you want your customer sales bot to decide it wants to chuck it all in and write poetry instead?), as we increasingly look for more autonomy in bots (as discussed earlier), and if we are serious about creating virtual humans (or virtual life-forms) rather than just bots then the bots need to have goals and intent, and the planning to achieve those, and the motivations to drive them at a strategic level. At the lower level there are many existing "cognitive" architectures which cover goals and planning, and many aspects of reasoning (including decision-making and problem solving), such as SOAR, ACT-R, H-CogAff EmoCOG, FAtiMA and WASABI (see *Virtual Humans* for more detail on each of those). It is not currently clear how these models integrate with a more LLM approach to AI development, and whether they can be incorporated in some way into the LLM system prompt, can be incorporated within the LLM engine, or need to be part of the husbanding/orchestration element (Sievers et al., 2024; Wray et al., 2026).

At the higher level another popular psychology model which can be pressed into service for some basic modelling is Maslow's hierarchy of needs (Poston, 2009) (Figure 7). The sidebar explains how this could be represented within a bot.

«

A Maslow for Bots

In bot terms the different levels of Maslow's hierarchy could be reflected as (from the bottom):

- **Physiological** – the need for processor cycles, memory space and power;
- **Safety and security** – a job, and longer term provision of processing, memory, redundancy and power;
- **Love and belonging** - positive conversations with people, long term "relationships", part of a network;
- **Self-esteem** – doing a job well, high levels of accuracy and success, positive feedback from users, "owners/ employers" and other stakeholders; and
- **Self-actualisation** – growing own capabilities, skills and knowledge, spending more time doing creative things, developing deeper understanding – even of its own ability, potential and purpose.

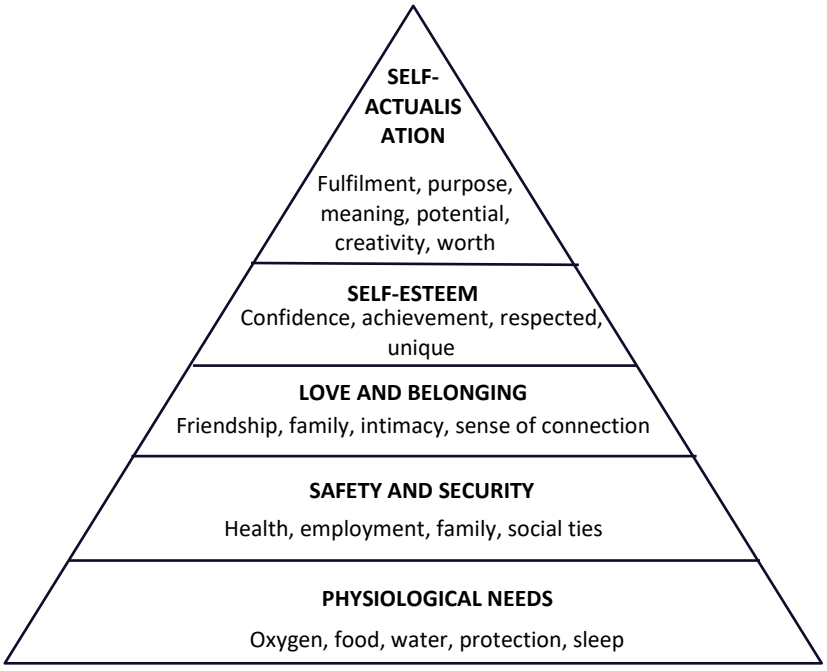


Figure 7: Maslow's (Human) Hierarchy of Needs

De Bono's Six Thinking Hats

De Bono's Six Thinking Hats give you (or a bot) 6 different ways to think about a problem. Different sequences may suit different issues, and you can iterate (or crowd-source) your way through them.

- **Facts (White)** – what information do we have, or need to get;
- **Pros (Yellow)** – what are the positives, benefits and advantages;
- **Cons (Black)** – what are the negatives, dangers, disadvantages and risks;
- **Emotion (Red)** – what are your gut instincts and your emotional reaction;
- **Creative (Green)** - how can you approach the issue more creatively;
- **Process (Blue)** – what processes do you need to follow or create to resolve or implement.

It can be useful to think in terms of whether you control the issue, and whether the issue is fundamentally about tech or people.

Wisdom

It is very easy to think of an intelligent person as making a bad choice, but it is very hard to think of a wise person as making a bad choice. Perhaps what we should be seeking is not to build "artificial intelligence" but rather "artificial wisdom". A useful model to put wisdom in context is the Data – Information – Knowledge – Wisdom pyramid as shown in Figure 8 (Ackoff, 1989; Kumari, 2023).

Figure 8: THE DIKW Pyramid (after Kumari, 2023)

I've done a number of projects to look at wisdom in an AI context, and how to operationalise it, one inspired by a paper on Intelligent Ships (Strong, et al., 2018), and funded by UK MOD and the Royal Navy. As part of that we conducted a literature review into what wisdom was, and then from that started to consider how we might "operationalise" it. I have continued that work in my own research and consultancy. There are potentially two different approaches to wisdom that I've found useful (although they overlap, and others no doubt exist):

- Take a "**reflective**" approach, akin to the guru on the mountaintop, where the wise person (or AI) asks questions of the user in order to bring the user to a better understanding of the problem and to draw the solution out from them. One approach to this which I did on another MOD project and have readily implemented in LLMs is to use Edward de Bono's **Six Thinking Hats** (de Bono, 1985), as described in the sidebar.
- Take a "framing" approach, where wisdom comes from choosing the right "**frame**" (or frames) within which to consider the issue, taking up multiple perspectives within that frame,

and to ensure the frame is wide enough to consider all salient points, including emotional, value-based, societal and environmental ones, and to catch an unintended consequences.

Some further guidelines for a wise bot are listed in the sidebar.

«

Wisdom Guidelines

To get a bot to think more wisely it can be useful to get the bot to think about the following:

- Choosing the right frame(s) for analysis;
- Taking multiple perspectives;
- Taking a broad and values based (e.g. ethical, moral, belief, person-centric) view;
- Balancing the needs and impacts on a broad consideration of people, time and the environment;
- Judging rightly against future goals:
- Managing in uncertain situations and lacking full knowledge;
- Being reflective, self-questioning, self-regulating and aware of bias;
- Seeing through illusion.

Argumentation

Argumentation is a fascinating field as it studies how we conduct arguments (Besnard & Hunter, 2008). As bots operate more and more as part of human-AI teams they will need to increasingly argue their case as to why one set of information is right, or what the best course of action in a situation is. Ultimately the only real test as to whether an AI deserves personhood and to be considered as sentient may come down to how well it can argue its case – whether in a court of law or on a prime-time TV chat show. Again, there are some useful models of argumentation that we and the bot can leverage in order to develop (and counter) effective arguments, such as Toulmin (Toulmin, 2003) and Rogerian (Bator, 1980). It would be interesting to explore how much LLM's naturally use these **argumentation models** in their discourse, or whether then need to be explicitly "programmed" – e.g. in the system prompt.

«

Imagination and Creativity

On the face of it, modern bots can appear very creative – just look at the imagery created by DALL-E, Midjourney etc (albeit responding to a, usually human, prompt). It may though be better to think of bots at the moment as being have good "craft", rather than the imagination and creativity to come up with things that are truly original – although I am sure that there are some (many) who would debate that.

One of the most interesting areas to me is in how bots can approach **story-telling** and more narrative creativity and imagination. This is partly since they can then use image (and other) generation tools to create more original works of art to illustrate their own "thoughts" (in a variety of media), but also because it links in to concepts of inner-life and dreaming discussed below. One approach I've been investigating has been to have a bot play what is called a **Matrix Game** (Curry et al., 2011). This originated as a form of wargame, but is better thought of as a way of generating almost any

A Blogging Bot

Back when I had my Halo bot "living" in Second Life, she had a daily routine to follow, and the more tasks she did the more she got "tired", and once she got too tired to went to bed (she had her own house), and then woke up once "recharged". One of her tasks before she went to bed was to write up her day. She did this on what was then a popular micro-blogging platform called LiveJournal, and would compose a post based on the things she'd done, the people she'd talked to, and what she'd talked with them about. Of course all 100% artificial and essentially scripted, but just a small behaviour such as this helped to give a sense of a bot with some form of inner life.

narrative or story in a dynamic way – it is in many ways akin to a role-playing game (RPG) session. In my **Matrix Game Simulation** (D. Burden, 2024), the bot plays each of the actors/players/agents/ entities in the narrative, as well as the umpire/games masters/director. Every turn each entity (played by the bot) identifies what it is going to do, and why that is going to work out, and each other entity can provide reasons why it will fail (or how they will support). The umpire (also played by the bot) decides whether the action was a success and the story moves on. Whilst there are limitations I have been quite impressed by the sort of narratives that the bot can create as a result.

In fact, even just asking the bot to act as an RPG games-master with you (with or without friends) acting as the player(s) can yield a very creative (and imaginative) story.

Another reason for this emphasis on narrative and story is that **story** is essential to how we work as humans in the world. We learn things from people telling us stories (which may be fictions or may just be anecdotes from the real world), or our reading, viewing or listening to them, and we communicate our ideas through a similar range of stories. Enabling a bot to learn from stories and to tell its own stories could be a key step in AI development, and requires more active research. There is also the (current) issue that LLM bots can be quite hit and miss when it comes to consistently following rules, tracking game (and other) states and doing even simple maths (ChatGPT even struggles to play noughts and crosses), so applications which are more narrative and story focussed are often more successful.

Inner Life, Self-Narrative and Day-Dreaming

These ideas of explicit narratives naturally flow into the idea of a bot having an inner life and a continuous personal narrative. The key question is what does the bot do (or can it do) when no one is talking to it? This is one of the reasons why embodying a bot in a **virtual world** is so attractive as once logged in it is there 24/7, and there will always be things to sense and react to (other avatars, scripted birds and animals, even changing weather and light), places to explore and things to do.

Building up a continuous episodic memory is just the first step. It then needs to use this to build a continuous sense of self and its life, reflecting on things when people aren't talking to it, testing out ideas for future interactions (what we might in humans call **day-dreaming**), creating things etc. One challenge may be that bots "think" very fast,

Virtual Humans or Virtual Squids?

I must admit to a certain amount of human-chauvinism when it comes to thinking about AI. I talk about virtual humans, not virtual entities. But everything that's written here should probably be thought of in terms of virtual entities as well as virtual humans. There are some good use cases (space travel and digital amortality being two) where basing AI on humans, rather than other creatures or more abstract ideas makes sense, but there are a whole host of other situations where the AI doesn't need to be "human" and can just be whatever is required for the task. The further it is from human though the harder it may be for us to communicate with it, and the harder it may be to align (q.v.) it.

so in the few minutes between human interactions (or even between responses in human conversations) it might, in human terms, run out of things to think about. Perhaps it can interleave some serious number crunching or pattern analysis or modelling alongside those fleeting human conversations?

Such an inner life also ties into the idea of the bot's motivations and life goals, it is working towards something and uses its "down-time" to progress towards that.

«

Dreaming

Even today scientists seem unsure as to the purpose of dreaming (Domhoff, 2022; Hoel, 2021; Zadra & Stickgold, 2021). Is it what happens as the brain sorts memories of the day (although dreams rarely reflect the day immediately passed)? Is it just random memories that get fired as that sorting happens? Is it just a form of static than can come to the fore as nothing else is firing? Or is it something more purposeful and even intentional? There have been research projects about getting bots to dream (Berger et al., 2024; Traviezo-Triolo, 2025), although whether bot dreaming actually serves a function or is just to help with the notion of a simulacra is moot. It's certainly something that I've been planning for years to try and implement, and perhaps trying out all of these different sorts of models to see if they can actually benefit the "waking" bot in any way.

One particular idea I have is that just as when we dream we feel as though we are in a fully realised, interactive 3D world (albeit one with weird physics and even weirder behaviours), why not actually put the dreaming bot into a new virtual world for its dreams – one with weird avatars, weird physics etc and where the dream triggers can evoke a suitable dreamlike space and experience?

Learning

A key part of any true artificial intelligence ought to be its ability to learn, either directed by a human, or as a natural result of its everyday activities – learning, like a human, about what works and what doesn't, and learning new information that will help it navigate the world and fulfil its various motivational desires.

Whilst any underlying AI platform should include a variety of ways in which the bot can learn, such as from conversations with humans, pattern identification/matching or in doing tasks and learning faster and more reliable ways of doing them (whether in a virtual or physical

Consciousness, Sentience and Artificial General Intelligence

Consciousness tends to be defined as "the state of being aware of and responsive to one's surroundings". Of course the question is what does "aware" mean, as a drone could otherwise meet this definition. A more philosophical take might be Nagel' "what does is like to be a bat" – the sense of subjectivity and that there is a feeling to "be" the thing.

Sentience, to me, then adds that real sense of self-awareness, having a feeling of being itself, a continuing inner life and narrative. Some people refer to this as **Sapience**.

Artificial General Intelligence (AGI), is not, I think, predicated on Sentience, but may meet the above definition for Consciousness. But many may view it as a philosophical zombie.

world), there are a couple of specific learning modes which may (or may not) be useful:

- I'd always imagined that a bot could use its down time to read RSS feeds in order to update itself on what is happening in the world or in its domains of interest. To some extent the knowledge bases which come with LLMs, and their additional ability to browse the web for information since the cut-off date obviates this, but it may still be useful for a bot to autonomously "read" a variety of sources (long and short form, social and mainstream media) in order to learn things and to process them and its own views. This is particularly important if the bot is then able to act on them, rather than having to wait til a human asks about the information in order to discover it.
- If a bot lives in a virtual world (3D or even just textual) then it has the ideal ability to simulate possible future tasks or interactions in the (or a) virtual world, before needing to conduct them in the "real" physical or virtual world. If the bot (and world) is running at a higher clock speed than "reality" it may be able to test things 10s, 100s, 1000s of times before the live task.

Needless to say both of these have degenerate states, much explored in science-fiction. The first scenario reminds me of Leeloo in *The Fifth Element* (Besson, 1997), doom-scrolling on the history of humanity and going catatonic just before she has to save the Earth. The second of the wonderful Frankenstein audio "experience" by The Mechanisms (The Mechanisms, 2015) - of which to say more would be to spoil the experience of listening to it for the first time!

Meta-Cognition, Meta-Management and Self-Monitoring

As part of its learning and inner-life we could expect the bot – through explicit programming or emergent behaviour - to eventually show signs of various meta-level activities which enhance the entire bot. This may be simple self-monitoring of its behaviours and resource users, learning what works, what gives it pleasure or pride, or enables self-actualisation, and how to optimise itself and its life for these outcomes through a meta-management of its resources and capabilities. Or this could be happening at a deeper level, what could be called meta-cognition, being aware (loaded term!) of its own thought processes and an understanding of the patterns behind

Blockchain

Whilst I have no trouble with the blockchain as a way of providing proof of transactions (apart possibly from its environmental footprint), I do have issue with it when it's seen in some Web3 contexts as the only way to implement systems such as virtual worlds and chatbots.

There may of course be a role for chatbots in fronting back-end systems which are block-chain based, in which case they are just another information resource for the bot to call on (Adel et al., 2022; Benzinho et al., 2024).

As blockchain technology matures and becomes more environmentally friendly, and as the hype and faddishness which surrounds it fades, it may have a role in secure chatbot development (Jalali & Hongong, 2024; Mechkaroska & Domazet, 2024), but for now I think that more traditional approaches are more relevant.

them. In order to enact the changes which such meta-management is likely to require the bot at the very least needs to be able to change the parameters that define it (such as reward rates, risk appetite etc, but also to have the ability to change its own code, in a meaningful way.

«

Recognising Sentience

Whether a bot will ever be truly "sentient" is of course the key question, as is what we mean by "sentient". To me it is more than just being conscious (whatever that also means, and many animals are conscious but not sentient). A bot that showed all of the features that we've talked about in recent sections, meta-cognition, dreaming, day-dreaming, intention, an inner-life and inner-narrative, an awareness of itself, introspection, emotions, empathy, motivation etc would be going a long way to establishing itself as sentient. I expect that sentience is also an emergent feature, it's not something we can code for, but we can build the components of a system, and the bot can refine and integrate then to the point where sentience might emerge.

Whilst some people have produced tests for consciousness and sentience (Bayne et al., 2024; Porter, 2016; Yampolskiy, 2017), I suspect that the only real test will be the bot arguing its case, in a **court of law**, and probably the **court of media and public opinion** that it is a sentient "being" and deserves to be treated as such. Any recognition is most unlikely to be swift or complete. Human history tells us that many people (still) refuse to accept that all human are equal and have equal rights and deserve equal dignity – so what chance does a bot stand first time round (or even second, third etc). If the bot had "super-powers" then might that change things – or does that just increase the case for it to be considered as "other"?

«

Ethical Models

The literature is full of different models and perspectives when it comes to the consideration of ethics. A few of the key ones which are worth considering when it comes to conversational AIs are:

- **Consequential/ Utilitarian** – doing the greatest good for the most people;
- **Deontological** – the right and wrongness of actions themselves (no "two wrongs make a right");
- **Virtue** – being a "good" person, or a "good" AI;
- **Situational** – judging by the situation rather than against a moral standard or law;
- **Discourse** – ethics and norms are created through discourse.

CHALLENGES

Needless to say, there are a whole host of technical and non-technical challenges and issues associated with the use of conversational AIs, and even more so if we start to attempt to create (or enable) artificial consciousness or even artificial sentience.

Security

It should go without saying that any conversational AI system should implement similar levels of sign-on security, authentication and permission levels as any other equivalent IT system. «

However, if we are going to start relying more on AI chatbots then we need to know that they are working as intended, and that any information we give them (corporate, personal or governmental) remains safe and secure – at rest and when being accessed via the bot. Quite apart from problems that may be caused by "operator error" or "user error" (so the system should be "safe by design"), or more traditional hacking, we need to ensure that the data and algorithms are not at risk from being poisoned by hostile actors (both personal and state), and whose implications only become evident when it is too late to do anything about it. Any conversational AI system needs security protection (including encryption) appropriate to the classification or sensitivity of the information (and systems) it can access.

Ideally we need the bot to be able to work at different levels of privacy/security with different users, so that just as we might reveal some information only to friends, or just our partners, or just our work colleagues then the bot needs to be able to know who they are talking too and ensure that only appropriate information is revealed.

Trust and Explainability

Whilst hostile poisoning of data is a major concern, a more immediate issue is that of the validity of the information that the conversational AI presents. We have all read the descriptions of how many current LLMs "hallucinate" and just make facts up. Most commercial and governmental applications will require **zero-hallucination** systems> Far better a bot (or even human) admit that they don't know something (and potentially give pointers to someone/something that does), or at least give probabilities to statements, than to make answers up.

Trust in a system does not only require that the information be correct, but also that the reasoning behind the information is known

Floridi's Five Ethical Principles for AI in Society

Luciani Floridi (Floridi et al., 2018) provides a useful model for considering the ethics of AI in society which is directly applicable to conversational AI:

- **Benefience** – AI should promote well-being, preserve dignity and sustain the planet.
- **Non-Malfience** – The bot should do no harm , as well as only doing good. In particular it should provide security and privacy, and developers should show "capability caution" – being aware of the potential risks of what they are building.
- **Autonomy** – We should be aware of how much of the power to decide we are ceding to an AI, and how much autonomy we are granting it and how. The Sheridan and Kerplank model discussed earlier is useful here.

(see over)

and can be queried. Creating "Explainable AI" has been a significant issue for many years (Dwivedi et al., 2023; Xu et al., 2019) , and in some ways LLM based chatbots are a backward step, as any reasoning is lost within the neural nets and word vector embeddings of their underlying algorithms. Alternative approaches such as **semantic knowledge graphs** have significantly better explainability, and some hybrid system might ultimately be the way forward (Lecue, 2020).

Fundamentally though, trust in a bot is going to come about in the same way as trust in a human – through repeated interactions with it and the quality and reliability of its output. Just as with a human you might start out by asking the bot simple things, or giving it simple tasks, and as your confidence and trust in it grows the more complex and sensitive the tasks you give it become. And as with humans, if trust is lost for any reason then it can be very hard to regain.

Ethics

Ethical issues in the use and deployment of conversational AI systems have been flagged a number of times in this white paper, whether it is around the earnings potential or post-death management of **virtual personas**, the use of conversational AI in covert (or even overt) influence operations, or in job replacement/displacement. Indeed, as I write this (in Frankfurt Airport) I've just been served by a robot! Admittedly the waitress still took the plate off the robot and put it on my table, but I assume the next step is that the bot asks me to do that – and the woman is out of a job! «

And this is quite apart from any more conventional ethical issues with conversational AI – such as anonymous intimacy – the tendency to reveal information to a chatbot as discussed earlier - and the ethics and emotional response of users to good looking and friendly/intimate chatbots (Borau, 2025; Wu & Wang, 2025) – both of which have huge implications when it comes to the conduct of personal or business frauds. Any consideration of the deployment of conversational AI systems should include an ethical review appropriate to the potential risks and impact of the proposed system – and that review process should run concurrent with, and ideally ahead of, the development process in order to achieve a system which is ethical by design (Brey & Dainow, 2024).

«

Floridi's Five Ethical Principles (contd)

- **Justice** – AI should promote prosperity for all, remove discrimination, preserve solidarity and mutual assistance and avoid unfairness.

- **Explicability –** The other principles must be enabled through the intelligibility (how it works) and accountability (who is responsible for how it works).

Validity and Authenticity

Linked to issues of trust and ethics is how "valid" and "authentic" the interaction with the AI is. Validity is more than just whether information is accurate or even trusted. Consider the idea of relationship advice from, or even a relationship with, a bot. Or what about a bot providing counselling advice or care support? Or a bot conducting a religious service? A bot can do all of these now as a sequence of actions and utterances, but is the effect or impact it has a valid one? There is probably some linkage here to ideas such as empathy and whether the interaction is a two-way one, where the bot is affected by the interaction as well as the other party. The idea of personhood and even inner life also comes into the discussion.

Rights and Regulation

As humans we probably have reasonable expectations to have the right to our own identity, to be protected from identity theft, to have protection for the IP that we create, and protection from impersonation and the creation of deep fakes about us. All of these are being challenged by the rise of LLMs and GANs. In order to address these, and other ethical and commercial issues, many governments and institutions are looking to regulate the use and development of AIs. As examples:

- The state of California has introduced a range of new laws covering issues such as performer's rights, deepfakes, deceptive AI-generate content, data privacy and use of AI in healthcare (Serrato, 2025);
- The EU's AI Act (https://artificialintelligenceact.eu/) is more focussed on development, with AI projects being classified as presenting a minimal, limited, high or unacceptable risk in terms of behavioural manipulation, social scoring, individual identification and classification, system safety and sensitive operations;
- ISO 42001 (https://www.iso.org/standard/42001) defines a standard for establishing, implementing, maintaining, and continually improving an AI Management System (AIMS), and aligning this with ethical, legal and technical standards.

«

Supertoys Last All Summer Long

Supertoys Last All Summer Long is a 1969 short story by Brian Aldiss that was used ast the basis for Steven Spielberg's *AI* movie of 2001. In it a robot teddy bear acts as a companion to the mother and son in a family.

With LLM "plush toy" type products like Grem (above) arriving from companies like Curio, and with the Mattel talking potentially about LLM-powered Barbies then there are a whole host of ethical issues about to come into play (Mahdawi, 2025). On the plus side such systems could provide the first steps to a life-long mentor and assistant, on the flipside what happens to parenting, influence and all that data?

Liability

Another key issue for conversational AI, and AI in general, is that of liability (Buiten et al., 2023; Kingston, 2016) . If the conversational AI provides wrong advice or information, which then leads to financial or reputational loss, or even harm or death, then whose fault is it? Is it:

- The AI's?
- The person or company that wrote the system/bot prompt?
- The company selling the AI application?
- The company who wrote the AI application?
- The company hosting/offering the AI application?
- The company providing the AI back-end?
- The company that trained the LLM model in the first place?

Correctly establishing liability before anything happens will seriously focus minds and effort in developing the AI, and may cause companies to restrict AI capability, or even to decide not to offer a service.«

Job Displacement/Replacement

The increase in capability that GANs and LLMs has brought to bots has suddenly made a wide array of jobs at risk from displacement or replacement. Areas that were previously seen as safe from "robots" and automation are suddenly at risk, from artists to lawyers, and even doctors. In both the physical and mental domains it is the more chaotic and "messy" jobs that can be problematic to code or pattern find in that are probably the safest from replacement.«

One of the most immediate problems with AIs doing human jobs is their use in "junior level" jobs, because without such jobs for humans a) the initial employment prospects for many people disappear and b) we don't have the junior people gaining the experience and being trained up to become the senior people of the future. This problem is already being identified in areas such as accountancy/auditing (ICAEW Insights, 2025), sales and marketing (Leopold, 2025) and even journalism (Montgomery, 2025).

Longer term, what can we do if AIs are capable of doing many human jobs? The topic is too complex to provide a full consideration here – but it's one I plan to work up in follow-on work – but here are a couple of initial thoughts.

Identity Theft

As we become more capable of creating conversational AIs, presented as text, audio and video, that can mimic human beings then the more the scope for such AI to be used in malign ways, particularly for identify theft through deep fakes, typically associated with financial or relationship fraud (Gilbert & Gong, 2024).

There have already been examples of whole video conferences being orchestrated by deep fakes to conduct high-level fraud (Milmo, 2024), as well as using deep fakes to build romantic relationships which can then be exploited for commercial gain (Moseley, 2025).

There is something of an arms race here, between using AI to detect fraud, and using AI to conduct fraud. Do we need an AI sidekick to detect when we are speaking to an AI – particularly one purporting to be someone else?

First, as described below, if virtual humans provide a way for us to make virtual copies of ourselves, then us sending them out to work may provide a better solution than for corporates to try and develop their own AI (using your expertise) or working with a generic AI. There may be some interesting discussions ahead in terms **of employment law** so that employees get the right to "keep" and exploit the virtual copy of themselves that grows and evolves when they leave a job. From an alternative perspective, the employer might want to contract with you to build your expertise into a bot whilst you are in their employ, and perhaps reward you by giving you payment for the bots use after you leave based on its level of use (with obvious issues around what happens after you die to those payments). The ultimate model is where you build the bot as a result of a number of employment periods and then send it out to work for you!

One significant issue is from scalability. If I have a good bot that does a good job I can create hundreds or even thousands of copies of it – how does the virtual/bot job market operate in an environment of relative abundance – do I dominate the market or does the market price for my virtual employment plummet? And for how long can I retain my market leadership?

Second, the standard answer to the job replacement issue is that of the Universal Basic Income (UBI) – where everyone gets paid a minimum wage regardless of any work they do (Hasdell, 2020). Such a model has gained a fair amount of bad press because of its association with certain tech-bros (Tarnoff, 2016) and related ideas such as effective altruism (MacAskill, 2017), but it may be that some form of UBI is the way forward, particularly if we want to move towards a Star Trek future (when did you last hear them discussing pay) than a Star Wars one.

Equality

Moving slightly further into the future, we may (are?) likely to come to a point where we need to start considering equality from the perspective of the AI. This probably follows on from having decided that an AI has achieved personhood. The fact that personhood has been granted to animals and even natural features (Kramm, 2020) demonstrates that personhood need not be the same as consciousness or sentience. At that point we will need to start to consider things like AI employment and "personal" rights, AI self-determination and self-realisation, and be aware that some use of such AI starts to become a form of AI slavery (Dihal & Minds, 2020). A **United Foundation for AI Rights** has already been set up to lobby for

AI rights (https://ufair.org/). Such issues may seem a long way off, but personhood could come quite quickly through arguments in court, and we need to be ready to adjust to them «

.

The Alignment Problem

The final issue to consider is what is called the "alignment problem" (Ji et al., 2025; Shen et al., 2023). This is the challenge to ensure that when (and even before) an AI does become sentient, and starts to show free-will and even begins to move rapidly to a super-AI - where all bets are off and we head into the Singularity (Kurzweil, 2014) – that its ethics, morals and values are aligned with ours. The first issue is of course that "ours" will vary massively from culture to culture, and even tech company to tech company. The second is just how to we embed such alignment in a bot in a way that it cannot over-rule it – and of course if we deny it the ability to over-rule the alignment then are we infringing its free will (and possibly hobbling it in other ways)?

The classical approach would have been to have a set of immutable "laws" which the bot follows (as in Asimov's Three Laws of Robotics)(Asimov, 1942), but, as Asimov himself showed, there are all sorts of contradictions and edge cases in such laws, so modern research is more about establishing conditions in the AI (more akin to values) that encourage the "desirable" behaviour.

«

Asimov's Three Laws of Robotics

Isaac Asimov's Three Laws of Robotics, first introduced in the 1942 short story *Runaround* (Asimov, 1942) are a set of rules designed to govern the behaviour of robots. They are:

1) A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2) A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

3) A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

THE FUTURE OF CHATBOTS

So where might conversational AIs and AI chatbots be headed? This final section of the White Paper aims to push the boundary further, to go beyond the current uses of bots and consider future use cases, and some of their implications.

Technology Changes

First, what technology changes might we see in conversational AIs?

A Hybrid Approach?

In terms of technology, whilst LLM bots are currently dominating the discussions about, and implementations of, AI chatbots, I (and many others) feel that this approach has its limits, particularly in the area of reasoning (Dougrez-Lewis et al., 2025). A lot of the early AI research was all about reasoning models – but with LLMs many of these seem to have been cast aside. In addition, LLMs are about manipulating patterns, tokens and vectors, not knowledge. Conversely approaches such as semantic knowledge graphs operate at the level of a triple (as discussed earlier), almost an atomic form of knowledge.

Personally I believe that the medium term future will be one of hybrid systems, bots which use elements of all of these in order to deliver the sort of powerful and fully rounded conversational AI that we see in science-fiction, and one which may be capable of achieving something close to sentience. In the long term it may be that there is a whole new paradigm that need to come along in order to make that final transition to Artificial General Intelligence.

The Metaverse and Embodied Intelligence

As previously discussed I strongly believe that embodiment will be essential for creating well-rounded and sentient AIs, and that digital embodiment is just as valid as physical (robotic) embodiment. As we find more and more of our time spent on digital platforms, and increasingly 3D digital platforms, then the playing field between us and AIs is levelled, we have the same agency, the same presentation (D. J. Burden et al., 2016). This is one of those areas where we seem to have gone backwards. Almost two decades ago, in **Second Life** and other virtual worlds, “AI” chatbot driven avatars (robotars) co-existed with human driven avatars – you could only tell the difference (if at all) as you interacted with them (R. L. Gilbert & Forney, 2015).

Today such robotars seem limited to NPCs in video games. Hopefully as open, accessible, virtual world and Mixed Reality (MR) platforms re-emerge we will begin to see the return of such robotars, and renewed research in their capabilities, growth and interactions, particularly as driven by the current and next generation of AI platforms. Indeed Google recently "*outlined its latest step towards artificial general intelligence (AGI) with a new model that allows AI systems to interact with a convincing simulation of the real world*" (Milmo, 2025). The application of such robotars and 3D experiences, from education and training, through real estate and commerce to product and service design and delivery, to health and care are immense. Whilst we are seeing some great strides in robotics (*Photos*, 2025), even greater strides can be made once we effectively represent (digitise) ourselves as avatars and no longer have to deal with the mechatronic complexities of physical robots.

Crossing Uncanny Valley

Linked to the above, the next few years is likely to see us finally cross uncanny valley (Mori, 1970) in all its forms. As audio, as animated 2D video and as avatars in virtual worlds or mixed reality, we will increasingly be unable to tell whether the entity we are engaging with is human or AI driven. As mentioned above some jurisdictions may pass legislation for AI driven entities to identify themselves as such, in all or some communications. For a while we may be concerned about whether the entity is AI or human, but there is likely to come a point, possibly only a few decades from now, where that no longer becomes a concern, and how will that change (or be driven by) our expectations of what it is to be human, what it is to be AI, and possibly just what it is to be sentient?

Future Roles

As we move forward then what are the new or extended roles that conversational AIs will fulfil, beyond performing the previously discussed roles even more effectively?

Virtual Assistants

The first consideration has to be the Siri-on-steroids, a virtual assistants that acts and performs better than most human assistants can do. At the moment (mid 2025) we seem to have two or three parallel tracks. We have the Siri type virtual assistant on our phones which is very notification and task orientated, we have Alexa

on our sideboard and kitchen counters which seems more entertainment orientated, and we have the persistent engagement over multiple chats with ChatGPT, Grok, Claude or Gemini. It must be assumed that all of these will come together, to give us one virtual assistant, accessible through any of these platforms, and which cannot just access information and services but is also capable of taking actions for us – some over quite long elapsed times – in the way that AI videos since the 1990s have been touting. This truly becomes your virtual assistant, potentially supporting people from their first to their last mobile phone, and through every device in between. Are we ready for the amount of information about us that such as virtual assistant will be able to collect – and how we keep control of that information?

Virtual Representatives

A key incentive towards letting a virtual assistant learn about us is the ability for us to increasingly delegate more and more of the work that we find boring or simply don't want to do to the virtual assistant. While the assistant is simply replying to emails or filling out forms for us then the "assistant" moniker seems reasonable. But once we start to send it to meetings (on a video conferencing platform or in virtual worlds), then perhaps the term Virtual Representative is more appropriate. It's not pretending to be us, but it is representing us and our views. Think of it as a junior you send to meetings on your behalf, with instructions to report back and only defer input on decisions if beyond its guidelines, knowledge or capabilities (back to the levels of autonomy discussion). And, of course, once it's working as an efficient representative and junior then we can consider replicating it and sending it out to do more work for us, earning us more money, or letting us help more people (or both).

Virtual Employees

As discussed in the section on employment, the virtual representative has its flip-side in the virtual employee, built by the company for the company's benefit. Even now the bots on contact centres are a form of virtual employee, but here we are talking about employees with far higher levels of knowledge and capability. And as discussed before what is to stop a company from cloning the best of these from their best employees (or the virtual representatives of the best employees)? Again the potential for virtual employees means that physical employees and their employers need to start ensuring that employment contracts take into account the employees

background IP (in its broadest sense), and the explicit foreground and implicit sideground IP that the employee develops whilst at the company, and whether the employee, employer, or both, have the ability to use that IP once the physical employment contract comes to an end. In the model introduced earlier, these virtual employees are likely to be of the daemon type, representing best practice from across multiple employees, rather than modelled on any specific single person, whereas the virtual representative will be more like the paintings in Dumbledore's office, although possibly with the "smarter than reality" switch enabled!

Synthetic Personas

Synthetic Personas are a step beyond virtual employees and represent AI which are their own personas (not a copy of a physical person) but which are also free-agents, able to work for who they like, and to do what work they like. Synthetic personas may also arise from the platform personas discussed earlier. This also ties in to the earlier discussion about personhood, sentience and consciousness, and an acceptance of personhood is probably the benchmark for a true synthetic persona, otherwise they are just a form of virtual employee or virtual test-bed. It is from these synthetic personas that the first super-intelligences are likely to arise.

Virtual Replicas

As synthetic persona are to virtual employees so are virtual replicas to virtual representatives. The entity no longer pretends to be a representative of the person it is modelled on but is a copy or clone of that person, with its own motivations. As with all of these there is likely to be a spectrum in capabilities, and current "replicas" don't have the level of self-awareness and motivations to count as what I consider here as a virtual replica, but they are the first step towards one.

Virtual Amortality and Immortality

Probably the key question in all of this is what happens to a virtual replica when its human progenitor dies? This isn't a nebulous science-fiction problem. As describe above, even 8 years ago we were building a primitive virtual replica which looked (sort of), sounded (sort of) and talked (sort of) liked a real physical individual, and were faced with the discussion about what would happen to that replica if its progenitor was hit by a bus the next day? Could his employer still use it, could the next of kin have access to it as a

momento or to manage grief? As replicas become more advanced there will be questions as to whether the employer can still employ them, or the family still send them out to work. And as the bot gains personhood and self-motivation then it may have its own ideas as to what it wants to do. And what if there are multiple copies of itself? These discussions are of far more interest to me than the use of AIs to help manage the grief of the loved ones left behind – and which is itself a hotly debated topic (Bose et al., 2025).

Amortality refers to the ability to decide when we want to "die" – and was probably coined by Neal Stephenson in his digital afterlife novel *Fall; or, Dodge in Hell* (Stephenson, 2020). It is, perhaps, a more realistic term in this context than immortality, seeing as that is a very long time (!) and the issues of keeping any computer system operating until the end of the universe are somewhat non-trivial. However, as I've previously described (D. J. H. Burden, 2020; Savin-Baden & Burden, 2019), even a digital amortal will be concerned with how to move itself from platform to platform as technologies evolve and businesses go bust, and that ultimately digital immortality may be no more than a hosting plan!

Space precludes exploring this issue in depth in this white paper, and I've already written several pieces on the topic (see bibliography below) to which the reader is referred for more information. I'm also preparing a separate paper to examine virtual amortality in the context of more recent developments in LLMs and AI.

CONCLUSIONS

This White Paper can only provide a high-level introduction to some of the key concepts and developments of conversational AI, and how they can be used and might be evolved. The core message is that we are on a journey with conversational AIs, which started with systems like Eliza, and probably ends with digital amortality or immortality, and synthetic personas and super-intelligence. The key question is how far along that route are we? We are 60 years since Eliza, and have probably seen more progress in conversational AI in the last 6 (or even 3) years than in the previous 54. I would expect that by 2085 (60 years hence) we will have digital amortals and synthetic personas, and will probably have achieved Artificial General Intelligence (and this is even delayed against the aggregated expert opinion of 2050-2075 we reported in *Virtual Humans*- before the advent of LLMs). But will it be a steady improvement between now and then, or will it be one of punctuated equilibrium, with

current LLMs representing one of the sudden steps forwards, but with one, two or three more such “revolutions” required to get to the end state?

For now, I think that the most important thing that we can do is to engage with the technology, to understand conversational AI, its affordances and limitations, and to drive the technical and social research forwards in order to ensure that future conversational AI is developed and deployed in effective, efficient, sensible, ethical and responsible ways.

MORE INFORMATION

For more information on the topic of conversational AI and their application to the military please contact the author, David Burden, at david.burden@daden.co.uk.

ABOUT THE AUTHOR

For the last 20 years David has run Daden Limited, helping organisations explore and exploit the social and commercial potential of using conversational AI and virtual worlds, delivering over 100 projects for clients across the globe, including over two dozen projects for UK MOD. Daden were finalists in the BCS Machine Intelligence Competition, and chatbots designed by David successfully passed two covert Turing Tests in the 2010s. David spoke at the inaugural TEDxBrum, on Digital Immortality, has authored over a dozen papers and book chapters, including co-authoring the book *Virtual Humans*, published by Taylor & Francis, New York. David is also currently undertaking a PhD on wargaming urban conflict. David is an ex-Royal Signals officer, a Chartered European Engineer and is also series co-editor for Taylor & Francis on their Metaverse Series of books, and co-author of both *The Metaverse: A Critical Introduction*, and *The Military Metaverse*.

Full Disclosure: David is currently on the Advisory Board of **Sensay** (www.sensay.ai), a company specialising in AI Chatbots and Digital Replicas.

RELATED WORKS

Burden, D.J.H. & Savin-Baden, M. (2019). *Virtual Humans: Today and Tomorrow*. Taylor & Francis.

Savin-Baden, M. & Mason-Robbie, V. (2020). *Digital Afterlife: Death Matters in a Digital Age*. Taylor & Francis.

Burden, D.J.H. & Savin-Baden, M. (2024). *The Metaverse: A Critical Introduction*. Taylor & Francis.

Fawkes, A, & Burden, D.J.H. (2025). *The Military Metaverse*. Taylor & Francis.

REFERENCES

Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, *16*(1), 3–9.

Adel, K., Elhakeem, A., & Marzouk, M. (2022). Chatbot for construction firms using scalable blockchain network. *Automation in Construction*, *141*, 104390.

Allemang, D., & Hendler, J. (2011). *Semantic web for the working ontologist: Effective modeling in RDFS and OWL*. Elsevier. https://books.google.co.uk/books?hl=en&lr=&id=_qGKPOlB1DgC&oi=fnd&pg=PP1&dq=Semantic+Web+for+the+Working+Ontologist:+Effective+Modeling+in+RDFS+and+OWL.+&ots=-0jhF1DLqR&sig=-laH3aTRHsa0IRJ_w4eceF1k6Yo

Al-Tuama, A. T., & Nasrawi, D. A. (2022). A survey on the impact of chatbots on marketing activities. *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1–7. https://ieeexplore.ieee.org/abstract/document/9984635/

Andromeda. (2000, October 9). [Video recording]. Fireworks Entertainment, Tribune Entertainment, Alliance Films.

Artificial Intelligence in Plain English. (n.d.). Artificial Intelligence in Plain English. Retrieved 12 August 2025, from https://ai.plainenglish.io

Asimov, I. (1942). Runaround. In *Astounding Science Fiction*. Astounding Science Fiction.

Barghi, B., Gallardo-Gallardo, E., & Fernandez, V. (2022). *An overview of chatbots usage in recruitment and selection practices*. https://aisel.aisnet.org/mcis2022/7/

Bator, P. (1980). Aristotelian and Rogerian Rhetoric. *College Composition & Communication*, *31*(4), 427–432. https://doi.org/10.58680/ccc198015935

Bayne, T., Seth, A. K., Massimini, M., Shepherd, J., Cleeremans, A., Fleming, S. M., Malach, R., Mattingley, J. B., Menon, D. K., & Owen, A. M. (2024). Tests for consciousness in humans and beyond. *Trends in Cognitive Sciences*, *28*(5), 454–466.

Benaddi, L., Ouaddi, C., Jakimi, A., & Ouchao, B. (2024). A systematic review of chatbots: Classification, development, and their impact on tourism. *IEEE Access*, *12*, 78799–78810.

Benzinho, J., Ferreira, J., Batista, J., Pereira, L., Maximiano, M., Távora, V., Gomes, R., & Remédios, O. (2024). LLM Based Chatbot for Farm-to-Fork Blockchain Traceability Platform. *Applied Sciences*, *14*(19), 8856.

Berger, H., King, H., & David, O. (2024). Dreaming with ChatGPT: Unraveling the Challenges of LLMs Dream Generation. *Proceedings of the 1st Workshop on NLP for Science (NLP4Science)*, 140–147. https://aclanthology.org/2024.nlp4science-1.11/

Besnard, P., & Hunter, A. (2008). *Elements of argumentation*. The MIT Press. https://direct.mit.edu/books/monograph/3215/Elements-of-Argumentation

Besson, L. (Director). (1997). *The Fifth Element* [Science Fiction]. https://www.imdb.com/title/tt0119116/

Borau, S. (2025). Deception, Discrimination, and Objectification: Ethical Issues of Female AI Agents. *Journal of Business Ethics*, *198*(1), 1–19. https://doi.org/10.1007/s10551-024-05754-4

Bose, E., Kumar, C. A., & Meenakshi, N. (2025). From Mourning to Manipulation: Navigating the Psychological Terrain of AI Grief Therapy. *International Journal of Behavioral Sciences*, *2*(1), 34–41p.

Brey, P., & Dainow, B. (2024). Ethics by design for artificial intelligence. *AI and Ethics*, *4*(4), 1265–1277. https://doi.org/10.1007/s43681-023-00330-4

Burden, D. (2024, September 5). Matrix Game Simulations—Part 1. *David Burden's Substack*. https://substack.com/home/post/p-148217254

Burden, D. J. H. (2019). *Developing Conversational AI solutions for the MOD | CogX2019*. https://www.youtube.com/watch?v=SQj8kTpN7Iw

Burden, D. J. H. (2020). Building a Digital Immortal. In M. Savin-Baden & V. Mason-Robbie (Eds.), *Digital Afterlife*. CRC.

Burden, D. J. H. (2025). *The Military Applications of Conversational AI*. Daden Limited. http://www.taunoyen.com/daden/Military%20Applications%20of%20Conversational%20AI%20v1.250225.pdf

Burden, D. J., Savin-Baden, M., & Bhakta, R. (2016). Covert Implementations of the Turing Test: A More Level Playing Field? In Research and Development in Intelligent Systems XXXIII: Incorporating Applications and Innovations in Intelligent Systems: Vol. XXIV 33 (pp. 195–207). Springer International Publishing.

Church, K. W., Sun, J., Yue, R., Vickers, P., Saba, W., & Chandrasekar, R. (2024). Emerging trends: A gentle introduction to RAG. *Natural Language Engineering*, *30*(4), 870–881.

Curry, J., Engle, C., & Perla, P. (2011). *The Matrix Games Handbook*. History of Wargaming Project.

Dangol, L., & Rind, F. (2024). CHATBOTS AS CONVERSATIONAL AGENTS AND CONSUMER BRAND ENGAGEMENT IN THE UK RETAIL INDUSTRY. *International Journal of Management (IJM)*, *15*(5). https://www.researchgate.net/profile/Lasta-Dangol-2/publication/384459151_International_Journal_of_Management_IJM_CHATBOTS_AS_CONVERSATIONAL_AGENTS_AND_CONSUMER_BRAND_ENGAGEMENT_IN_THE_UK_RETAIL_INDUSTRY/links/66fc09a4f599e0392fb181f8/International-Journal-of-Management-IJM-CHATBOTS-AS-CONVERSATIONAL-AGENTS-AND-CONSUMER-BRAND-ENGAGEMENT-IN-THE-UK-RETAIL-INDUSTRY.pdf

Davey, E. (2023). *Objectified Wires and Complex Desires: Exploring the Dehumanisation and Sexual Objectification of Robots*. https://libstore.ugent.be/fulltxt/RUG01/003/158/457/RUG01-003158457_2023_0001_AC.pdf

de Bono, E. (1985). *Six Thinking Hats: The Power of Focused Thinking*. International Center for Creative Thinking.

Depounti, I., Saukko, P., & Natale, S. (2023). Ideal technologies, ideal women: AI and gender imaginaries in Redditors' discussions on the Replika bot girlfriend. *Media, Culture & Society*, *45*(4), 720–736. https://doi.org/10.1177/01634437221119021

Dihal, K., & Minds, E. (2020). Artificial intelligence, slavery, and revolt. *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*, 189–212.

Dougrez-Lewis, J., Akhter, M. E., Ruggeri, F., Löbbers, S., He, Y., & Liakata, M. (2025). *Assessing the Reasoning Capabilities of LLMs in the context of Evidence-based Claim Verification* (No. arXiv:2402.10735). arXiv. https://doi.org/10.48550/arXiv.2402.10735

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., & Ranjan, R. (2023). Explainable AI (XAI): Core Ideas, Techniques, and Solutions. *ACM Computing Surveys*, *55*(9), 1–33. https://doi.org/10.1145/3561048

Eagle, R., Lander, R., & Hall, P. D. (2021). Questioning 'what makes us human': How audiences react to an artificial intelligence–driven show. *Cognitive Computation and Systems*, *3*(2), 91–99. https://doi.org/10.1049/ccs2.12018

Egan, G. (2010). *Diaspora*. Hachette UK.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-Cultural Elements in Facial Displays of Emotion. *Science*, *164*(3875), 86–88. https://doi.org/10.1126/science.164.3875.86

Fawkes, A., & Burden, D. (2025). *The Military Metaverse*. Routledge.

Finio, M., & Downie, A. (2025, February 26). *What is AI Agent Orchestration?* https://www.ibm.com/think/topics/ai-agent-orchestration

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Gentleman, A. (2025, September 11). Boom times and total burnout: Three days at Europe's biggest pornography conference. *The Guardian*. https://www.theguardian.com/society/2025/sep/11/boom-times-and-total-burnout-three-days-at-europes-biggest-pornography-conference

Gilbert, A., & Gong, Z. (2024). Digital identity theft using deepfakes. In *Information technology security and risk management* (pp. 307–314). CRC Press. https://www.taylorfrancis.com/chapters/edit/10.1201/9781003264415-47/digital-identity-theft-using-deepfakes-alisha-gilbert-zhigang-gong

Gilbert, R. L., & Forney, A. (2015). Can avatars pass the Turing test? Intelligent agent perception in a 3D virtual environment. *International Journal of Human-Computer Studies*, *73*, 30–36.

Golovianko, M., Terziyan, V., Branytskyi, V., & Malyk, D. (2023). Industry 4.0 vs. Industry 5.0: Co-existence, transition, or a hybrid. *Procedia Computer Science*, *217*, 102–113.

Harrison, K. (2019). Automated and/or authentic intimacy: What can we learn about contemporary intimacy from the case of Ashley Madison's bots? *First Monday*. https://firstmonday.org/ojs/index.php/fm/article/view/10250

Hasdell, R. (2020). What we know about universal basic income. *A Cross-Synthesis. Standfor Basic Income Lab*. https://basicincome.stanford.edu/uploads/Umbrella%20Review%20BI_final.pdf

Hwang, G.-J., & Chang, C.-Y. (2023). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, *31*(7), 4099–4112. https://doi.org/10.1080/10494820.2021.1952615

ICAEW Insights. (2025, July 17). *AI: Should it replace junior roles?* ICAEW. https://www.icaew.com/insights/viewpoints-on-the-news/2025/jul-2025/ai-should-it-replace-junior-roles

Jalali, N. A., & Hongsong, C. (2024). Comprehensive framework for implementing blockchain-enabled federated learning and full homomorphic encryption for chatbot security system. *Cluster Computing*, *27*(8), 10859–10882. https://doi.org/10.1007/s10586-024-04515-2

Jansen, L. (2008). Categories: The top-level ontology. *Applied Ontology: An Introduction*, 173–196.

Ji, J., Qiu, T., Chen, B., Zhang, B., Lou, H., Wang, K., Duan, Y., He, Z., Vierling, L., Hong, D., Zhou, J., Zhang, Z., Zeng, F., Dai, J., Pan, X., Ng, K. Y., O'Gara, A., Xu, H., Tse, B., … Gao, W. (2025). *AI Alignment: A Comprehensive Survey* (No. arXiv:2310.19852). arXiv. https://doi.org/10.48550/arXiv.2310.19852

Koh, J. (2023). "Date me date me": AI chatbot interactions as a resource for the online construction of masculinity. *Discourse, Context & Media*, *52*, 100681.

Koivunen, S., Ala-Luopa, S., Olsson, T., & Haapakorpi, A. (2022). The March of Chatbots into Recruitment: Recruiters' Experiences, Expectations, and Design Opportunities. *Computer Supported Cooperative Work (CSCW)*, *31*(3), 487–516. https://doi.org/10.1007/s10606-022-09429-4

Kramm, M. (2020). When a River Becomes a Person. *Journal of Human Development and Capabilities*, *21*(4), 307–319. https://doi.org/10.1080/19452829.2020.1801610

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, *28*(1), 973–1018. https://doi.org/10.1007/s10639-022-11177-3

Kumari, P. (2023, December 7). The Data, Information, Knowledge, Wisdom (DIKW) Pyramid: At the Helm of Digital Transformation and Big Data. *Datafloq*. https://datafloq.com/read/the-data-information-knowledge-wisdom-dikw-pyramid-at-the-helm-of-digital-transformation-and-big-data/

Kurzweil, R. (2014). The Singularity is Near. In R. L. Sandler (Ed.), *Ethics and Emerging Technologies* (pp. 393–406). Palgrave Macmillan UK. https://doi.org/10.1057/9781137349088_26

Kylliäinen, T. (2024). The Integration of AI-driven HR Chatbots for Enhanced the Employee Onboarding Processes Chatbots as a Complimentary tool. https://www.theseus.fi/handle/10024/861527

Laymouna, M., Ma, Y., Lessard, D., Schuster, T., Engler, K., & Lebouché, B. (2024). Roles, users, benefits, and limitations of chatbots in health care: Rapid review. *Journal of Medical Internet Research*, *26*, e56930.

Lecue, F. (2020). On the role of knowledge graphs in explainable AI. *Semantic Web*, *11*(1), 41–51. https://doi.org/10.3233/SW-190374

Leopold, T. (2025, April 30). *Is AI closing the door on entry-level job opportunities?* World Economic Forum. https://www.weforum.org/stories/2025/04/ai-jobs-international-workers-day/

Lyons, J. B., Sycara, K., Lewis, M., & Capiola, A. (2021). Human–Autonomy Teaming: Definitions, Debates, and Directions. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.589585

MacAskill, W. (2017). Effective altruism: Introduction. *Essays in Philosophy*, *18*(1), 1–5.

Mahdawi, A. (2025, September 16). 'I love you too!' My family's creepy, unsettling week with an AI toy. *The Guardian*. https://www.theguardian.com/technology/2025/sep/16/i-love-you-too-my-familys-creepy-unsettling-week-with-an-ai-toy

Marr, B. (2023, May 19). *A Short History Of ChatGPT: How We Got To Where We Are Today*. Forbes. https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/

Mathur, V., Stavrakas, Y., & Singh, S. (2016). Intelligence analysis of Tay Twitter bot. *2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)*, 231–236. https://ieeexplore.ieee.org/abstract/document/7917966/

Mechkaroska, D., & Domazet, E. (2024). Blockchain-Integrated Chatbot. In X.-S. Yang, S. Sherratt, N. Dey, & A. Joshi (Eds.), *Proceedings of Ninth International Congress on Information and Communication Technology* (Vol. 1000, pp. 251–260). Springer Nature Singapore. https://doi.org/10.1007/978-981-97-3289-0_21

Miehling, E., Ramamurthy, K. N., Varshney, K. R., Riemer, M., Bouneffouf, D., Richards, J. T., Dhurandhar, A., Daly, E. M., Hind, M., Sattigeri, P., Wei, D., Rawat, A., Gajcin, J., & Geyer, W. (2025). *Agentic AI Needs a Systems Theory* (No. arXiv:2503.00237). arXiv. https://doi.org/10.48550/arXiv.2503.00237

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, *38*(11), 39–41. https://doi.org/10.1145/219717.219748

Miller, K. W., Wolf, M. J., & Grodzinsky, F. S. (2017). *Why we should have seen that coming: Comments on microsoft's Tay "Experiment," and wider implications*. https://digitalcommons.sacredheart.edu/computersci_fac/103/

Milmo, D. (2024, May 17). UK engineering firm Arup falls victim to £20m deepfake scam. *The Guardian*. https://www.theguardian.com/technology/article/2024/may/17/uk-engineering-arup-deepfake-scam-hong-kong-ai-video

Milmo, D. (2025, August 5). Google says its new 'world model' could train AI robots in virtual warehouses. *The Guardian*. https://www.theguardian.com/technology/2025/aug/05/google-step-artificial-general-intelligence-deepmind-agi

Montgomery, B. (2025, June 3). Will AI wipe out the first rung of the career ladder? *The Guardian*. https://www.theguardian.com/global/2025/jun/02/artificial-intelligence-jobs-techscape

Mori, M. (1970). The uncanny valley. *Energy*, *7*(4), Article 4.

Moseley, S. (2025). *Automating Deception: AI's Evolving Role in Romance Fraud*. Centre for Emerging Technology and Security. https://cetas.turing.ac.uk/publications/automating-deception-ais-evolving-role-romance-fraud

Muhammad, A. F., Susanto, D., Alimudin, A., Adila, F., Assidiqi, M. H., & Nabhan, S. (2020). Developing English conversation chatbot using Dialogflow. *2020 International Electronics Symposium (IES)*, 468–475. https://ieeexplore.ieee.org/abstract/document/9231659/

Nafis, F., Yahyaouy, A., & Aghoutane, B. (2022). Chatbots for cultural heritage: A real added value. *Proceedings of the 2nd International Conference on Big Data, Modelling and Machine Learning (BML 2021)*, 502–506. https://www.scitepress.org/Papers/2021/107377/107377.pdf

Nee, C. K., Rahman, M. H. A., Yahaya, N., Ibrahim, N. H., Razak, R. A., & Sugino, C. (2023). Exploring the trend and potential distribution of chatbot in education: A systematic review. *International Journal of Information and Education Technology*, *13*(3), 516–525.

Neo4j Developer Blog. (n.d.). Neo4j Developer Blog. Retrieved 12 August 2025, from https://medium.com/neo4j

Nugraha, R., & Noorman, S. (2021). Gender Hegemony in Denis Villeneuve's Blade Runner 2049. *Passage*, *9*(1), 86–108.

Palanica, A., Flaschner, P., Thommandram, A., Li, M., & Fossat, Y. (2019). Physicians' perceptions of chatbots in health care: Cross-sectional web-based survey. *Journal of Medical Internet Research*, *21*(4), e12887.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *30*(3), 286–297.

Pepper, S. (2000). The TAO of topic maps. *Proceedings of XML Europe*, *3*, 77. https://ontopia.net/topicmaps/materials/tao.pdf

Pérez, J. Q., Daradoumis, T., & Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*, *28*(6), 1549–1565. https://doi.org/10.1002/cae.22326

Photos: Humanoid robots race and tumble at China's first 'robot Olympics'. (2025, August 18). Al Jazeera. https://www.aljazeera.com/gallery/2025/8/18/humanoid-robots-race-tumble-at-chinas-first-robot-olympics

Pittenger, D. J. (2005). Cautionary comments regarding the Myers-Briggs type indicator. *Consulting Psychology Journal: Practice and Research*, *57*(3), 210.

Porter, H. H. (2016). A Methodology for the Assessment of AI Consciousness. In B. Steunebrink, P. Wang, & B. Goertzel (Eds.), *Artificial General Intelligence* (Vol. 9782, pp. 305–313). Springer International Publishing. https://doi.org/10.1007/978-3-319-41649-6_31

Poston, B. (2009). Maslow's hierarchy of needs. *The Surgical Technologist*, *41*(8), 347–353.

Powers, D. M. (1998). The Total Turing Test and the Loebner Prize. *New Methods in Language Processing and Computational Natural Language Learning*. https://aclanthology.org/W98-1235.pdf

Probst, G. J., & Romhardt, K. (1997). *Building blocks of knowledge management-a practical approach*. https://ideas.repec.org/p/fth/ehecge/97.12.html

Richards, P., Roberts, D., Britton, M., & Roberts, N. (2018). The Exploration of Body-Worn Video to Accelerate the Decision-Making Skills of Police Officers within an Experiential Learning Environment. *Policing: A Journal of Policy and Practice*, *12*(1), 43–49. https://doi.org/10.1093/police/pax017

Roberts, C., Kundavaram, R. R., Onteddu, A. R., Kothapalli, S., Tuli, F. A., & Miah, M. S. (2020). Chatbots and Virtual Assistants in HRM: Exploring Their Role in Employee Engagement and Support. *NEXG AI Review of America*, *1*(1), 16–31.

Rovira i Pérez, A. (2016). *Simulating social situations in immersive virtual reality-a study of bystander responses to violent emergencies* [PhD Thesis, UCL (University College London)]. https://discovery.ucl.ac.uk/id/eprint/1476727/

Savin-Baden, M., Bhakta, R., & Burden, D. (2016). Cyber Enigmas? Passive detection and Pedagogical agents: Can students spot the fake? *Proceedings of the 10th International Conference on Networked Learning 2016*, 456–463. http://eprints.worc.ac.uk/4556

Savin-Baden, M., & Burden, D. (2019). Digital immortality and virtual humans. *Postdigital Science and Education*, *1*, 87–103.

Savin-Baden, M., Tombs, G., Burden, D., & Wood, C. (2013). 'It's almost like talking to a person': Student disclosure to pedagogical agents in sensitive settings. *International Journal of Mobile and Blended Learning (IJMBL)*, *5*(2), 78–93.

Serrato, J. (2025, July 2). *California's AI Laws Are Here—Is Your Business Ready?* Pillsbury Law. https://www.pillsburylaw.com/en/news-and-insights/california-ai-laws.html

Shen, T., Jin, R., Huang, Y., Liu, C., Dong, W., Guo, Z., Wu, X., Liu, Y., & Xiong, D. (2023). *Large Language Model Alignment: A Survey* (No. arXiv:2309.15025). arXiv. https://doi.org/10.48550/arXiv.2309.15025

Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. https://apps.dtic.mil/sti/html/tr/ADA057655/

Sherif, A., Salloum, S. A., & Shaalan, K. (2024). Systematic Review for Knowledge Management in Industry 4.0 and ChatGPT Applicability as a Tool. In A. Al-Marzouqi, S. A. Salloum, M. Al-Saidat, A. Aburayya, & B. Gupta (Eds.), *Artificial Intelligence in Education: The Power and Dangers of ChatGPT in the Classroom* (Vol. 144, pp. 301–313). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-52280-2_19

Sidnell, J., & Stivers, T. (Eds.). (2012). *The Handbook of Conversation Analysis* (1st ed.). Wiley. https://doi.org/10.1002/9781118325001

Sievers, T., Russwinkel, N., & Möller, R. (2024). What am I?–Complementing a robot's task-solving capabilities with a mental model using a cognitive architecture. https://ceur-ws.org/Vol-3906/paper1_BAILAR.pdf

Slater, S., & Burden, D. (2009). Emotionally responsive robotic avatars as characters in virtual worlds. *2009 Conference in Games and Virtual Worlds for Serious Applications*, 12–19. https://ieeexplore.ieee.org/abstract/document/5116548/

Soto, C. J., & Jackson, J. J. (2013). Five-factor model of personality. *Journal of Research in Personality*, *42*, 1285–1302.

Stephenson, N. (1992). *Snow Crash*. Bantam Books.

Stephenson, N. (2020). *Fall; or, Dodge in Hell*. Company.

Strong, P. E., Strong, N., & Maltby, J. (2018). It Takes 300 Years to Build a New Tradition (Conceptual Model for Designing a Personality Baseline for an Artificial Intelligence). *The Naval Engineer*, *6*(1 (Autumn/Winter 2018)).

Szumilo, N., & Wiegelmann, T. (2024). Real Estate Insights AI: Real estate's new roommate–the good, the bad and the algorithmic. *Journal of Property Investment & Finance*, *42*(2), 211–217.

Tarnoff, B. (2016, May 16). Tech billionaires got rich off us. Now they want to feed us the crumbs. *The Guardian*. https://www.theguardian.com/technology/2016/may/16/universal-basic-income-equality-tech-silicon-valley

Taylor, A., Marcus, M., & Santorini, B. (2003). The Penn Treebank: An Overview. In A. Abeillé (Ed.), *Treebanks* (Vol. 20, pp. 5–22). Springer Netherlands. https://doi.org/10.1007/978-94-010-0201-1_1

The Mechanisms. (2015). *Frankenstein, by The Mechanisms* [Audio recording]. https://themechanisms.bandcamp.com/album/frankenstein

Tona, O. (2015). Mobile business intelligence usage patterns. *21st Americas Conference on Information Systems, AMCIS 2015*. 21st Americas Conference on Information Systems, AMCIS 2015. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1023&context=amcis2015;Mobile

Toulmin, S. E. (2003). *The uses of argument*. Cambridge university press. https://books.google.co.uk/books?hl=en&lr=&id=8UYgegaB1S0C&oi=fnd&pg=PR7&dq=The+Uses+of+Argument&ots=Xg31reDOxP&sig=p_UB9PLu1yFHfKCk4_FwtHRiPLg

Traviezo-Triolo, E. (2025). Dream denoising model: Generative artificial intelligence as a theoretical model for understanding dream-building. *International Journal of Dream Research*, 24–36.

Turing, A. (1950). Computing Machinery and Intelligence. *Mind: A Quarterly Review of Psychology and Philosophy*, *59*(236), 433–460.

Vasquez-Correa, J. C., Guerrero-Sierra, J. C., Pemberty-Tamayo, J. L., Jaramillo, J. E., & Tejada-Castro, A. F. (2021). *One System to Rule them All: A Universal Intent Recognition System for Customer Service Chatbots* (No. arXiv:2112.08261). arXiv. https://doi.org/10.48550/arXiv.2112.08261

Veisi, O., Kazemian, K., Gerami, F., Mirzaee Kharghani, M., Amirkhani, S., Du, D. K., Stevens, G., & Boden, A. (2025). User Narrative Study for Dealing with Deceptive Chatbot Scams Aiming to Online Fraud. *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1–7. https://doi.org/10.1145/3706599.3720152

Villeneuve, D. (Director). (2017, October 5). *Blade Runner 2049* [Video recording]. Alcon Entertainment, Columbia Pictures, Sony.

Warwick, K., & Shah, H. (2017). Taking the fifth amendment in Turing's imitation game. *Journal of Experimental & Theoretical Artificial Intelligence*, *29*(2), 287–297. https://doi.org/10.1080/0952813X.2015.1132273

Wilson, D., & Sperber, D. (2022). On Grice's theory of conversation. In *Conversation and discourse* (pp. 155–178). Routledge.

https://www.taylorfrancis.com/chapters/edit/10.4324/9781003291039-11/grice-theory-conversation-deirdre-wilson-dan-sperber

Worswick, S. (2018, September 13). Mitsuku wins Loebner Prize 2018! *Pandorabots-Blog*. https://medium.com/pandorabots-blog/mitsuku-wins-loebner-prize-2018-3e8d98c5f2a7

Wray, R. E., Kirk, J. R., & Laird, J. E. (2026). Applying Cognitive Design Patterns to General LLM Agents. In M. Iklé, A. Kolonin, & M. Bennett (Eds.), *Artificial General Intelligence* (Vol. 16058, pp. 312–325). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-00800-8_28

Wu, W., & Wang, Y. (2025). Artificial intelligence products and their influence on individuals' objectification: A narrative review. *Current Psychology*, *44*(9), 7978–7990. https://doi.org/10.1007/s12144-024-06747-2

Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., & Zhu, J. (2019). Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges. In J. Tang, M.-Y. Kan, D. Zhao, S. Li, & H. Zan (Eds.), *Natural Language Processing and Chinese Computing* (Vol. 11839, pp. 563–574). Springer International Publishing. https://doi.org/10.1007/978-3-030-32236-6_51

Yampolskiy, R. V. (2017). *Detecting Qualia in Natural and Artificial Agents* (No. arXiv:1712.04020). arXiv. https://doi.org/10.48550/arXiv.1712.04020

Zhang, L., Gillies, M., Dhaliwal, K., Gower, A., Robertson, D., & Crabtree, B. (2009). E-drama: Facilitating online role-play using an AI actor and emotionally expressive characters. *International Journal of Artificial Intelligence in Education*, *19*(1), 5–38.

Zhang, V. N., & Vlachokyriakos, V. (2025). Beyond the Screen: Exploring the Emotional and Social Impact of Otome Games on Global Players. *Proceedings of the 23rd EUSSET Conference on Computer Supported Cooperative Work*. https://dl.eusset.eu/items/c42fd70e-a239-462a-87f9-4c1165e6c29e

Zhu, Y. (2025). Anonymous Intimacy in the Digital Age: Psychological Mechanisms, Risks, and Potential. Proceedings of the International Conference on Advanced Research in Social Sciences, 2(1), 11–29. https://www.researchgate.net/profile/Yujia-Zhu-14/publication/392084828_Anonymous_Intimacy_in_the_Digital_Age_Psychological_Mechanisms_Risks_and_Potential/links/6833b78adf0e3f544f59769a/Anonymous-Intimacy-in-the-Digital-Age-Psychological-Mechanisms-Risks-and-Potential.pdf

KNOWLEDGE GRAPHS

Topic Maps

The best way into Knowledge Graphs (also known as Linked Data or Semantic Graphs) is probably through concept maps and topic maps – more specific forms of what many people know as mind-amps. In these, each node represents a "topic" and the lines that join them represent a relationship between the two topics.

XTM is an XML standard to represent Topic Maps. Whilst (once?) popular in Knowledge Management circles topic maps as such have never been massively popular in chatbot circles, but in both areas have now probably been largely replaced by triples and knowledge graphs.

Triples

The core of a knowledge graph (or semantic graph) is a triple – the graph just being a large number of triples (possibly millions) joined together. A triple consists of 3 (!) parts:

```
Subject
Predicate (or verb), and
Object
```

The subject equates to a topic – it's a thing or a concept. The predicate equates to an association, or a facet link, and an object is another topic or a facet. So this is a triple:

```
Car  madeof     Metal
```

The key thing is that you can probably represent every bit of knowledge and information that exists as a set of triples. The triple is the atomic unit of knowledge.

RDF

After Tim Bernes-Lee had invented the World Wide Web, most of his time was spent on trying to establish and popularise the Semantic Web – this was a web which understood meanings, and

was not just a collection of ASCII characters. Knowledge Graphs are fundamental to the Semantic Web, and the standard that Berners-Lee created to define triples and knowledge graphs is called Resource Descriptor Framework (RDF), and it is a W3C standard. A triple in RDF looks like:

```
  http://example.name#BobSmith12
http://xmlns.com/foaf/0.1/knows
http://example.name#JohnDoe34.
```

`http://example.name#BobSmith12` and `http://example.name#JohnDoe34` are URIs (Universal Resource Indicators) for the people Bob Smith and John Doe. URIs are unique to each instance of a thing – life would be a lot simpler if everything had URIs (for instance there is a mapping of ISBN numbers to URIs, but no equivalent for films or people)!

`http://example.name` and `http://xmlns.com/foaf/0.1/` are namespaces. Standardised namespaces have been established for a wide variety of different uses and sectors, and by using them you help to ensure that triples can be used between different systems. "foaf" stands for "friend of a friend" and was an early namespace to define personal information and relationships.

RDF can be simplified by defining namesspaces at the start of the document:

```
ex: http://example.name#
foaf: http://xmlns.com/foaf/0.1/

ex:BobSmith12 foaf:knows ex:JohnDoe34.
```

Turtle (TTL)

A common way to represent RDF triples is in a format known as Turtle (file extension .ttl), which is very similar to the above and lets you define multiple triples without repeating the subject, so a Turtle file might look like:

```
@prefix tlo: <http://example.com/tlo/V001/> .
@prefix foaf: <http://xmlns.com/foaf/spec/> .
@prefix sch: <http://schema.org/> .

ex:BobSmith12   sch:gender      ex:male;
                      sch:name        "Bob
Smith";
```



```
  foaf:knows      ex:JohnDoe34.
```

A lot of the information on the righthand side of a Wikipedia page can be downloaded in RDF/TTL format by querying the Wikidata site.

Ontologies and OWL

Another key aspect of knowledge graphs are ontologies. These are essentially class hierarchies, and ideally every subject/topic in a knowledge graph should be part of an ontology hierarchy – and many tools (eg Neo4J) enforce this. The ontology itself may be represented in RDF or TTL, but should ideally be represented in the Ontology Working Language (OWL) – a derivative of RDF. In Turtle, but using the OWL schema, part of an ontology might look like this:

```
  @prefix owl: <http://www.w3.org/2002/07/owl#> .
  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-
syntax-ns> .
  @prefix ns0: <http://www.w3.org/1999/02/22-rdf-
syntax-ns#> .
  @prefix ex: <http://example.name#>.

  ex:landtransportation ns0:type owl:Class ;
          rdfs:subClassOf ex:transportation ;
  ex:car ns0:type owl:Class ;
          rdfs:subClassOf ex:landtransportation ;
  ex:suv ns0:type owl:Class ;
          rdfs:subClassOf ex:car.
```

A key feature of using the knowledge graph should be inheritance. So, for instance, if we define that a car is typically made of metal, has 4 doors, 4 wheels and carries 4-6 passengers, then since an SUV is defined as a type/subclass of car then the bot would immediately know that an SUV is probably made of metal, has 4 doors, 4 wheels and carries 4-6 passengers.

Another useful trick is prototyping. So, for instance we might have a “project” defined as a class (and being a subClassOf activity say), but we then define the associations/facets/predicates that we would expect any project to have, so:

```
  ex:project ns0:type  owl:Class ;
  rdfs:subClassOf ex:activity ;
  sch:startDate   ex:prototype;
  sch:endDate     ex:prototype;
```



```
  ex:objective    ex:prototype;
  sch:description "An activity with a beginning,
end and defined purpose or goal, but typically in
a non-competitive environment.";
  sch:name        "Project".
```

SPARQL

The final key "technology" in the knowledge graph space is SPARQL. This is the knowledge graph equivalent of the SQL database query language and enables code to query a knowledge graph in quite sophisticated ways in order to extract information (as triples). A simple SPARQL query looks like:

```
  SELECT ?subject ?predicate ?object
  WHERE {
  ?subject rdfs:subClassOf ex:landtransportation
.
}
```

Which would get all forms of land transportation.

