A picnic in the uncanny valley

or

How I learned to stop worrying and love AI. [1]

A report on the creation of a dynamic AI self-portrait, June-October 2024

Draft 6, Simon Penny, 10 Feb 2025

Abstract

The goal of *AI self-portrait* was to realise a persuasive simulation of a specific person in a specific professional mode, using AI tools and techniques of mid 2024. As a is a media-critical artwork, the purpose of the exercise was to explore the dimensions and underlying assumptions of the technology, by pursuing project that pushes the technology to its limits in certain ways. The creation of AI generated life-like characters is a topic in popular culture, and a technical goal in contemporary AI research, one that has huge commercial potential; at the same time is of great concern to writers and actors (as witnessed by recent Hollywood strikes). As an artwork, the project utilises irony and double coding. It is presented as a bona-fide attempt to generate an AI simulated academic, devoid of 'art' framing. It is interesting *because* it is boring.

This paper documents the development of an AI art project pursued according to a practice-driven-research methodology. The project is a response, not only to the rapid commercialisation of machine learning-based LLM and media/graphical tools, but also to the double-edged public discourse of dire dystopian warnings (that, nonetheless, drive interest, any publicity being good publicity in the 'AI gold-rush') and huge commercial success. The concerns of public intellectuals contrasts with a seeming supplication by the administrations of educational institutions, belying a failure of criticality in exactly the locations where one ought to expect it.

Keywords

AI-art, critical-art, LLM, machine learning, digital-twin, artificial characters.

1. Motivation, History and project goals

Over recent years, new AI tools have made deepfakes and automatically generated texts increasingly easy, the press has been full of utopian and dystopian narratives, and venture capitalists have thrown billions at AI startups in the 'AI goldrush'. The internet has seen a flood of trite 'AI-art' - garish scenes strongly reminiscent of the entire history of adolescent fantasy imagery since the mid C20th. I wondered what kind of critical art project might plumb the dimensions of the new AI and the rhetoric swirling around it. I thought an appropriately double-edged project would be to try to make a simulation of the academic me - to see just how close I could get to making myself redundant. This is, in a sense, 'doing the devil's work' (as one colleague observed, if successful, the university might be very

[1] *(subtitle adapted from the title of Stanley Kubrick's darkly satirical 1964 film *Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb*.)

interested). The project AI self-portrait, more colloquially 'Professor Simsimon' is a specialised self portrait of the 'professorial' me, analogous to 'expert systems' of the symbolic AI era. A reflexively critical goal was to test the capabilities of LLMs with respect to idiosyncrasy. We also wanted to test built-in guardrails by asking the LLM to take a critical of LLMs and the larger phenomenon of machine learning AI. I gathered a small group of collaborators and we began work in late May 2024. [2]

We began with the goal to create online academic presentations that looks like me, sounds like me and says things I might say, but haven't. Combining LLMs, voice cloning and deepfake video, we attempted to build an AI replica of the 'academic' me, delivering papers I might deliver. The LLM was trained on all my academic writing of the last 20 years. It was prompted to write academic papers, on topics within my range of processional writing - often containing critiques of AI and AI rhetoric. These papers were converted to transcriptions of spoken presentations (as it were). Voice-cloning was used to generate a facsimile of my speaking voice, and deepfake video was used to generate video imagery of me (talking head) speaking the voice clone. The LLM was used to generate slides for the talk and the whole thing was composited in simple video post-production.

This paper is written from the perspective of over 30 years as a practitioner in critical media arts, with long-term engagement in both technical and critical discourses in computing, digital cultures, AI and Artificial Life (see Penny 2017). The paper discusses the project in four parts. It offers a discussion of topics in the critical orbit of the project concerning AI, interventionist art practices and the larger socio-technical discourse. It then offers some reflections on the technical task itself, followed by a rough synopsis of the development process.

2. Theoretical discussion topics

Dynamic portraits – deepfakes, digital doubles and doppelgangers.
Deepfakes are presenting a crisis in the entertainment industry – broadly construed. At the same time, creation of 'believable' synthetic characters is a potentially highly profitable area and is being pursued as such. In popular culture, well-known cases of the technological double are Max Headroom (pic) and the Stepford Wives. South Korean popular entertainment is populated with virtual pop stars.[3] Alarmingly, purveying virtual romantic partners seems to be a viable business - the trend has worrying social implications.[4]

---

[2] My collaborators - all affiliates of UCI: Ge (Tom) Gao (junior ICS), Yiyang (Roger) Min (staff), Kenneth Pat (masters ICS), Yurun Song. (PhD candidate ICS)
[3] See https://kprofiles.com/ais-in-kpop-industry/ for an overview, (accessed 6jan25). There is also – hardly surprising – AI revenge-porn and porn deepfakes using the likenesses of known celebrities.
[4] The website for https://solcandy.ai/ (accessed 5jan25) begins "Generate your own Girl . Your dream companion awaits! Generate your AI girl, shape her look, personality, and bring her to life in one click. 100% powered by Artifical Intelligence." Hilariously 'artificial' is mispelled – should have used GPT😉

Historically, performative representations – portraits that simulate not only static physical appearance but cognitive qualities, gestures and vocalisations are a fixture in sci-fi - any number of cyborgs and 'persuasively' humanoid robots – from Maria in Fritz Lang's Metropolis, to Terminator. These serve a specific literary function, representing dystopic implications of mechanization by personifying philosophical crises regarding the relationship between humans and increasingly capable machines. In 'media art', Stelarc's Prosthetic Head (2014) was an early example of a digital portrait (pic). Avatar-based social media (Second Life, etc.) gave rise to (self)portrait avatars, and media artist Antoinette LaFarge pushed the notion of a social proxy into the real-world with her project World-Integrated Social Proxy (WISP).[5]

Reflexivity in critical media art

My art practice, for most of my career, has been located at the leading edge of technology development, critiquing the technology and the rhetoric around it by building critical artworks using (and often developing) the technology (see ~~Petit Mal, Fugitive, Traces~~). The members of the team – all computer scientists with no background in art, media art or critical theory, had no experience in this kind of work and approached the project in a pragmatic 'can-do' fashion, typical of engineering disciplines. Contrarily, my approach is guided by what Philip Agre called 'critical technical practice' (Agre 1997).

In developing Simsimon, we are pushing LLM / AI / Machine Learning against-the-grain in this sense – LLMs have a tendency to 'regress to the mean', that is, to find the most popular, and by implication the most ordinary 'answer'. But we are attempting to simulate the idiosyncracies of a specific individual. My approach in making technological interventionist research-driven art-practice has always had this disruptive edge - trying to make it do things it doesn't 'want' to do, seeing how it breaks, and what compromises have to be made. In effect, testing the paradigm by building things.

Irony and double coding – fakes and activism

The Yes Men, famous in the activist art world, have been wildly successful (as well as extraordinarily courageous, demonically clever and darkly funny) pursuing their projects in the public arena, never foregrounding the 'art' status of their interventions. Elaborate hoaxes are well known in the art world – notably, for instance, the faux archeology of Beauvais Lyons (in the eponymous Hokes archive) [6]; the fake biology of Louis Bec [7]; or the faux industrial history of Bonk in Finland [8]; the remarkable transgenic fantasies of Patricia Piccinini [9]. Antoinette LaFarge, in her magisterial "Sting in the Tale – Art, Hoax and Provocation" documents a long history of fakes and hoaxes in art and literature.

---

[5] The W.I.S.P. project debuted at the 2009 Digital Arts and Culture Conference at the University of California, Irvine (Director Simon Penny), where the W.I.S.P. project debuted. To prepare for this event, the actor Laura Kachergus worked closely with LaFarge and theater director Robert Allen.

[6] https://counterfactualartarchive.com/hoakes-archives/

[7] http://www.colloquebioart.org/pages/lbec.html

[8] https://www.bonkbusinessinc.com/, https://bonkcentre.fi/en/businessjananchovy/

[9] https://www.roslynoxley9.com.au/artist/patricia-piccinini

Simsimon – a parody or a hoax?
At the outset, Simsimon said ludicrous or meaningless things in ways that were unlike me. We laughed about how early 'talk' versions had me sounding like a 30 something hipster tech-bro giving a motivational talk (yecch!). Our original intention was to train thje LLM only on my work. In practice we found that it was impossible to wall-off the LLM from the larger internet. Even if the *content* was proscribed, GPT resorted to other resources on matters of style, language etc (hence the repugnant tech-bro persona). But as our techniques, and the technology, developed, Simsimon became unnervingly persuasive. In the process of refining our methods, we met new theoretical/aesthetic challenges (as well as new technical challenges). In the process of development of the AI self-portrait, a subtle aesthetic choice presented itself – was this to be a parody or a hoax? The difference being whether you let the audience in on the joke. We decided to play it straight, in order to exploit the anxiety of the viewer as a critical-activist technique. Perhaps, as a more literary commentator noted, the result is of the form of a *homage* or *pastiche*, which lacks both the criticality of parody or the intentional deception of hoax.

Representation, truthiness and meaning.
Like a self-portrait, and like all things resident in computers and on the internet, Simsimon is a representation, operating within the gamut of digital capabilities. At root - manipulating alphanumeric strings and generating outputs within a specific range of output capabilities of digital media such as text, video imagery and digital sound. That is, it is not materially existent - except as charges in electric circuitry. It does not occupy space in the same way I do, it is not alive - it does not metabolise food, and it cannot punch you. It does not 'experience the world' and crucially, it does not 'know' anything. Nor, I hazard, will it ever write anything like the above of its own volition (because it has no volition) - and because it is unlikely to be self-reflective, unless of course, instructed to be so, in which case it will emulate the style of someone being self-reflective. There is no 'there' there. It is all 'like' not 'is'. Everything seems and seems.

Simsimon is a fake – but what makes a fake persuasive? We are reminded of Steven Colbert's notion of 'truthiness'. The likeness is thin, but lacks disruptions that might disrupt an unwitting suspension of disbelief. It is after all a simulation of a representation of me encapsulated in a now-conventional media form, the streamed video lecture.

The LLM can assemble concepts I might assemble, using vocabulary I might use, in a persuasive presentation, but below the veneer of verbiage, there is nothing, no argument, because there is no motivating *attitude*. A gossamer tissue floating over a void, held aloft by warm breath - not even, just the simulation of warm breath. Like Cinderella's castle at Disneyland - its isn't a castle, it's a second order representation of a fantasy idea of a castle - a literary construction only. Baudrillardian precession of simulacra.

This brings us to a philosophical worry that has dogged AI for 50 years or more – the problem of *meaning*. Fundamentally, LLM cannot make 'meaning' because it doesn't

‘know’ anything. Sounding like me does not imply making sense like me. This *is* a philosophical question: we find it difficult to say what it means to ‘know something’, even for people. This begs another question – if it looks like me and sounds like me, but does not ‘make sense’, it is a likeness, perhaps, but is it a ‘self-portrait’? (building a babbling pseudo-me might have its own rewards).

Simsimon is not Simon
A novel aesthetic decision-point was reached – inasmuch as we were setting systems up to generate the best ‘me’ we were able, the ‘me’ we made had to stand on its own, and we had to refrain from editing and refining. In the same way that a portrait is not the person who is depicted, we had to let Simsimon be a different person from Simon – bearing a family resemblance to be sure - but if Simsimon said something that is consistent with the content and style of things Simsimon might say, but I did not feel that I would say that, I had to allow that Simsimon was allowed to say that because ‘he’ isn’t ‘me’. (To say ‘he took on a life of his own’ would be too dramatic).

While the technology is complex, it is now relatively easy to produce a moving video image of my face speaking, with of-the-shelf tools. If you know me well, after a few second you think - ok, something’s off. Likewise, the voice-clone. But it you did not know me, or had met me once years ago, you might interpret that oddness as media glitches - as we so often have learned to do, with bad video connections and so on.

Getting the LLM to ‘have ideas like mine’, and express them in the way I would, transpired to be far more difficult, if not impossible. This is due to a technical limitation of ML LLM in the form we had access to. GPT can discuss concepts I might discuss, using language I might use, yet there is an emptiness of content. Using the kind of language I might use is a superficial level under which lies the matter of meaning-making. What does it mean to have purchase on an idea? What is incisive reasoning? Whatever it is, we know what it is when we see it. LLMs apparently do not have that capability – at present. But it fakes it well enough that one has to be vigilant. This recognition demanded the application of scrupulous editorial analysis: ‘it sounds right, but is it actually saying anything (rational, rigorous and/or novel)?’

3. The socio-technical context

What is AI?
It is important to ensure meaningful discourse by clarifying many of the generalisations that infect (often-heated) public rhetoric. The term, like so many technical terms that enter in public discourse, is now applied to such a vast range of techniques and applications that it is nigh meaningless. Importqnt to note that what we now call AI is entirely unlike first generations symbolic AI of the 60s-90s. This is due to three developments:
• the rise of machine learning – that arose out of the tradition of neural network research (that was rejected by symbolic AI).

• the vast increase in speed and quantity of grunt processing (cf Nvidia).
• the existence of the internet and specifically, vast accumulations of datapoints in datacenters (server-farms)
As Michael Mateas succinctly summed it up a decade ago, symbolic AI applied complex algorithms to comparatively tine datasets, machine-learning AI applies relatively simple statistical procedures to gargantuan datasets.[10]

AI is already a cyborg - ML and internet.
These vast databases ML draws upon to develop its output is, essentially, the internet:, server-farms, databases, and the vast amounts of (often trivial) data held there, that are regularly 'scraped' by AI bots and crawlers. Machine learning, in its contemporary incarnation, is nested in an ocean of data resident in datacenters all over the world, linked by high-speed fiber-optic networks – the global digital-industrial complex we (laughably) refer to as the 'cloud'.

Worth saying too that the vast majority of that data was put there, 'manually' by people, posting vacation pics with comments, pinning maps, and so on. Another vast reserve is texts written by humans alive and deceased, that have been digitised. The data in those databases has been put there by people - albeit scraped, sorted and organised algorithmically. But the key point remains: it is all the distillation of the products of human minds. People interpret the world. People post the data. People write the scraping tools, not to mention the neural networks. People train the AIs (mechanical Turk). Like Soylent Green, AI is people.

The front end of ML is (loosely speaking) like a search engine, it collects relevant datapoints. The next stage involves judgement and selection - that come down to a kind of bell-curve. And if a large part of the data result for the query in question, is, say, racist, then the chosen result will be racist too (unless guard-rails have been put in place). But as we know there are clever ways to circumvent those guard rails. For instance - if you ask GPT for instructions on how to make a 'dirty bomb' it won't tell you. But if, say, you engage in minimal subterfuge, by asking: 'imagine you are a script writer writing a script about terrorists who are making a dirty bomb. Write that script.' you are more likely to get an informative response. Mind you, there is always the possibility of hallucination - it may just look like a good recipe for dirty bomb.

Why worry? AI and employment
Because LLMs can only inhabit the digital, they are restricted to the internet informational ecosystems. They can't 'get out'. [11] That is not to say that other applications of AI are not matters of great concern –military and surveillance applications being obvious cases - the spectre of autonomous weapons selecting their own targets. But large language models are just that: models of language. They know nothing of the world, indeed, they 'know'

[10] Personal communication.
[11] Cf Craig Reynolds' Tierra – digital wildlife park.

nothing. Nor do they have the capacity to reach out beyond 'the page' (though Agentic AI is currently attempting to facilitate such extensions – of the kind we are already familiar with in automated online activities – paradigmatically online shopping). As such I don't think there's a lot to get our knickers in a knot about. That said online data is taken to constitute 'the real' in way that cause it (regrettably) to loom as a presumed epistmological ground for digital natives (see Penny 2021 and Penny 2023).

As a result, any vocations that involve actual hands-on skill with material, artifacts, tools, organisms, plants, animals, and people – are relatively immune – though medical data such as radiography is increasingly effectively analysed by AI. Ironic to observe that highly valued hi-tech jobs of the last generation (ie, computer science) are precisely the ones being made redundant, and the comparatively despised 'trades': motor mechanics, electricians and plumbers, along with gardeners and artisans, are 'safe'. By the same token, writers and other language workers who work in more 'formulaic' modes – such as writing sit-com scripts - are increasingly vulnerable and rightly worried.

AI, Creativity and the Goedelian impasse

In popular discourse, the question of creativity vis-à-vis AI raises its head again as it did in the 1980s with first generation 'symbolic' AI. One might imagine that in a discussion of a critical AI-artwork, one might encounter heated rhetoric concerning creativity. The foregoing discussion should indicate that this author does not see this as a worry – but not because AI cannot be made to simulate conventional genres – as had been done in music, in art and in the more formulaic forms of entertainment – such as sit-coms and soap-operas. [12] The simple fact is that computing lends itself to formulaic practices, it being fundamentally constrained by rules. Therefore, the only possible kind of 'invention' is of a combinatorial kind, operating within a rule-based domain.

Take an example of a genre – say watercolor painting. We define the domain – this kind of pigment, these kinds of brushes, those kinds of surfaces – thing made within those constraints are deemed watercolor painting. Given these rules, an AI system might be able to paint any and all possible watercolor paintings – but it will never attach a candy wrapper or a bus ticket to the surface of the painting, because such acts are not within the domain. If we define the rules for sonnets or baroque fugues, systems will, and have, generate instances that conform. But we won't get quarter-tones or free-verse. The problem is that virtually all interesting creative acts either break the rules or make new ones. That is what we value as innovation, in the sciences as well as the arts. That is, such creative acts break the Goedelian frame. A loose paraphrase of Goedel's incompletenesss theorem is that one cannot describe a system with the rules of that system. That is, once cannot describe the rules of chess with the rules of chess, one has to work within a larger logical domain in

[12] In music, notably the simulations of J S Bach by David Cope, and earlier simulations of Miro 'constellation' paintings by Russell and Joan Kirsh, in the 80s, using LISP shape-grammars. See https://www.jstor.org/stable/40072590 (accessed 5jan25)

order to describe chess as, for instance, a variant of the class of board games. Unexpected but meaningful variety is not something computers can generate. No doubt a suitable AI image system could produce endless variations on say, Monet's water lilies, or (much more easily) Keith Haring graphics.[13] As radical scientific change involves paradigm shift (Kuhn), so truly innovative art (in the avant-gardist tradition) involves the reconfiguration of the conceptual framing and axiomatic assumptions of a practice. Computer code is built in a domain of logical rules, machine learning exists within that domain. To the extent that ML reference existing examples deemed (by some prior categorization) to be within the gamut, nothing can result that is outside that gamut.

Eccentricity and creativity – regression to the mean

Ironically, or perhaps predictably, Professor Simsimon is boring. While it is wryly amusing that the project is successful *because it is boring*, it does reveal an apparent limit of machine learning/large language models - eccentricity appears to be beyond its grasp, precisely because it's methods is to draw upon thousands of examples and will inevitably 'regress to the mean'. Any AI query draws upon millions of datapoints, and by some statistical processes loosely comparable to building a bell-curve, chooses something like the 'mean'. The underlying assumption is that the most common answer is *likely to be* the correct one. ML by its nature can only provide the most generic responses - mealy-mouthed mediocrity - because it is looking for the top of the bell curve on a million datapoints – its results can be nothing except ordinary.

This explains why such systems, by definition, *cannot be creative*, in the conventional sense. Creativity is, by definition, not 'ordinary,' it is not 'predictable', it is *unusual*. A key implication for our project is that idiosyncracy is *anathema to* such ML procedures. This reflects a deep lesson that is as relevant to educational policy as it is to AI about the cultural value of idiosyncracy and eccentricity. Statistically, the 'interesting' stuff is among the outliers - by definition, they're unusual, non-conforming, divergent and disruptive thinkers. [14]

4. Technical topics

Commercially available tools

At the outset, the project began with a typical computer-science research approach, the intention to prototype new tools. In the contemporary AI environment (as opposed to the artist-coder and hardware-hacker context the early days of digital media art) it became quickly clear that the technological complexities were so vast that a 'garage' approach was not viable. Instead, we were forced to adopt a more contemporary approach of researching and kluging together available online tools, many of them offered, temporarily, free, as is

[13] This apparently hasn't been done, but its low-hanging fruit.

[14] The product of AI will always be ordinary in this sense. Anecdotally, a colleague reported that using generative graphics tools in a class had the effect of raising the quality of the work of the 'lower half' of the class, but reducing the quality of the work of the better students. QED.

the pattern for emerging commercial software that simultaneously builds a user-base and exploits the experience and expertise of early-adopters, in the manner of non-commercial 'open-source' software development projects. This approach permitted more rapid development while limiting creative freedom, due to the closed nature of the tools, and their formulation towards envisaged needs of user-groups.

Not press-n-play, but second-order hands-on.
The projects discussed here were the result of concerted effort over nearly half a year by five specialists – including graduates and PhD candidates in computer science specialising in AI. Given that the goal of the project was to see how close the technology could get to a persuasive simulation of me, we approached the task with the goal to be as 'hands-off' as possible – not reaching-in and 'manually' adjusting for glitches etc.

Working with LLMs must be a disconcerting experience for engineers because there are no metaphorical levers to pull or buttons to press. I do not mean this disrespectfully, engineers are used to working with variables that have predictable and relatively direct effects – computer engineering is, after all (or was), *engineering*. Machine-learning is not amenable to the manual coding of old-school programming. The process is much more inferential – entirely unlike traditional explicit coding. With LLMs, one cannot 'find a bug' and rewrite.

With ML AI, you nudge and cajole, you poke it and see how it reacts, because it is, like all ML, inherently a 'black box': you can't reverse engineer it! The process of prompting - involving second-guessing how the AI will interpret certain instructions - feels more like gardening or training a puppy - hence the use of the term 'training' – a 'dance of agency' in Andrew Pickering's terms. The upshot was that we had to strategise ways to attract the LLM towards the kinds of expressions we were aiming at – allowing the systems to produce what they could produce, with limited editing and interference. Substantial time and effort went into refining prompts, refining the structures of prompts, and nudging the various tools towards a more accurate representation.

Synopsis of technical development

The development of these projects entailed numerous stages. None of this was 'push-button'. Each of these stages entailed substantial discussion, decision making, delegation and technical problem solving, that took the five of five months to achieve. This work was done by the AI Self-portrait team:-(whose ongoing efforts and collaboration is deeply appreciated).

Stage 1. uploading and processing all my professional writing for the last 20 years.
Stage 2. Designing workflow and testing tools.
Stage 3. devising and refining prompts to generate new texts on new topics

Stage 4. Reviewing text output to iteratively refine prompting and training – the problem of 'style'.
Stage 5. Voice-clone tool, trained on my voice, produced a simulation of my voice speaking the texts.
Stage 6. 'Deep-fake' video tool animated a still-frame of my face
Stage 7. slides were generated for the lecture, driving Latex from GPT
Stage 8. diagrammatic images were generated to illustrate the slides. Interfacing GPT with DALL-E
Stage 9. Use of GPT 4.o.1 introduced new literary capabilities and challenges for review of texts.
Stage 10. Multilingual Simsimon.

Stage 1 was to uploading and processing all my professional writing for the last 20 years. This included a 500 page book, a book manuscript in process, and dozens of published papers, along with numerous videos of lectures I've given. (Processing involved stripping away all titles, references, footnotes, page numbers and other ancillary text leaving only raw-text 'content'.
The plan was that this database would permit the LLM to say 'Simon-like things' about 'Simon-like topics'. This plan turned out to problematic, for two reasons – the database was not big enough, and it was difficult to 'wall-off' the LLM from consulting other online sources, which had the effect of reducing the 'Simon-like' nature of output material. This led to a range of strategies for making the output more 'Simon-like' – see below.

Stage 2. Designing workflow and testing tools. Originally our intention was directed at technical research, but fairly quickly, we shifted to a more pragmatic goal of combining available mostly free online tools, and thus also availing ourselves of online processing. This has the inevitable limitation that free tools don't stay free, and proprietors can change tools or remove them with minimal recourse or notice. Equally problematic, the product is a 'black-box' with only certain kinds of controls available on the interface surface. So as usual in tech art - for my career anyway, there's nothing 'timeless' in this project, it is tightly bound to the calendar of technological change. I'm well aware that if what we are doing now is concerning now, we may regard it as trivial next week.

We tested a number of LLMs (including GPT, LAMA, Gemini, etc) and settled on GPT, for a variety of fairly complex reasons to do with the way it draws on sources and flexibility of prompting and training. We chose Elevenlabs for voice cloning and Hedra for deepfake video generation.

As we tested prompting strategies, we realized that our initial prompts implied operations that involved multiple processes for the LLM. We found we got better performance by breaking-out prompts into smaller tasks, applying reductionism to good effect. This led to a strategy of sequential or nested prompts. Roger proposed an 'AI agent' architecture in which rather that treat GPT as one entity, we spawn-off multiple AI-agents, (replicating the

way we delegated work within our research team) - each of which was responsible for particular kinds of subtasks, and would pass things to each other.

Stage 3
Stage 3 of the process was devising prompts to generate new texts on new topics, that 'sound like me'. Initially there were two stages to this process – generating (short) academic 'written' papers, and then generating the equivalent of transcripts of spoken texts, that the voice-clone tool could then speak. A major challenge was getting the LLM to construct text that produced coherent arguments about recognizable subjects. In some cases, it was capable of formulating paragraph-length arguments, but subsequent paragraphs would take up different topics, with little connection.

A central difficulty has been the matter of 'style' - in choice of topics, vocabulary, sentence structure, phrasing, accent, vocal mannerisms, etc. LLMs have significant limitations in these areas. In my academic work, I use a complex vocabulary that combines terminology from numerous disciplines, including neuroscience, computing, cognitive science, philosophy, anthropology, the arts. A simulation of me has to build arguments that are like arguments I might make, using my vocabulary, in sentence constructions I would use.

The original intention was to put the LLM 'in a box' - to constrain the LLM so it only looked at my own work. This was already a limitation on the functioning of the LLM because it 'needs' orders-of-magnitude more data-points in order to function well – so we might induce 'hallucinations' by restricting it (which would in itself be interesting). Personally, I wanted to see the ways it failed under such conditions. The other team members became interested in achieving a contemporary version of a successful Turing test.

Stage 4 – review of generated texts.
In stage4, we encountered complex questions of personal style. When training ChatGPT to produce texts that sound like me (ie reads like a transcription of my spoken style), these characteristics of LLMs revealed themselves in amusing ways. At first, the text read like a 30-something hipster tech-bro giving a motivational talk or a TED talk. There was a Polly-Anna-ish *jollying* of the 'listener' that I, frankly, found infantilising. A supplicant and apologetic voice seems to be the default in these systems, belying a key commercial role, replacing call-center workers.

If Professor Simsimon is to be a successful simulation of me, it has to exhibit the curmudgeonliness of a disaffected senior professor, no longer attempting to pad his resume or play the career game, it has to have some wry double-edged humor, and a take perverse pleasure in upending stereotypes and disrupting polite norms. LLMs cannot do this – perhaps we should be grateful. As a friend who saw one of the videos – and was initially fooled - told me, he thought I'd drunk the academic kool-aid and had become a boring old fart. LLMs seem unwilling to be vehement, derisive, or satirical, they are polite, supplicant and annoyingly equivocal, presumably because they is made to do the work of call-center workers.

The LLM has to draw on diverse resources (such as language rules) to be able to, for instance, construct a sentence, arrange sentences in a sequence that 'follows', or build a comprehensible argument over several paragraphs. It has tendency to fall back on formulaic constructions. Apparently, according to these rules, a 'spoken presentation' has to include a vapid introductory passage about how excited the speaker is to be presenting this material today; and a concluding paragraph about what a pleasure it has been to bring this exciting topic to this audience and how I look forward to further discussions – none of which I would ever say, perhaps because I am a curmudgeon, or perhaps because the socio-intellectual milieu I inhabit grants its audience a level of maturity and interest that does not imply the need for 'motivation'.

Finding ways to discourage the LLM from falling back on generic formulae was a significant challenge – best achieved by crafting prompts to induce the LLM on a certain path, for instance "write a short spoken presentation on 'xyx' *as if it was given by* a 60-something year old professor educated in Australia".

Intros and outros – a tale told by an idiot.
It is a tale told by an idiot, full of sound and fury, signifying nothing. (Macbeth Act5 Sc5).

In one example of paraphrasing my written work as spoken presentation. In the body of the document, LLM did manage to sound like me, (using words I might use, in constructions I might make). But LLM made this (particularly egregious) summation:

So, as we reflect on these ideas, let's appreciate the lively interconnections between biology, cognition, and the arts. Each performance we witness, each story told on stage, offers us a window not just into human creativity but also into the very nature of life itself, however metaphorically we interpret it. Thank you for joining me today in this exploration.

I found this kind of output revolting. What is remarkable about this passage (and especially the middle sentence) is that it 'sounds right' but is fundamentally an empty collection of platitudes. Maybe that kind of 'word-salad' is a characteristic of a lot of human speech, but its not, I hope, of 'academic speech'. This has been the problem with generating the texts for Simsimon - so often it 'looks like' a ~~Simon Penny~~ paper, a ~~Simon Penny~~ idea, the right kinds of words are there in the right kind of order, but when analysed, little of consequence is being said - like so many political speeches - full of bluster perhaps, but with minimal substance. In this respect GPT is a *rhetoric machine*.

Stage 5
Eleven labs voice-clone tool was then trained on my voice and produced a simulation of my voice speaking the texts. On the acoustic level, *style* involves accent, intonation, emphasis, pacing and so on. At the lexical and grammatical level it involves sentence structure and vocabulary. The former is apparently relatively tractable – voice-clone tools can produce an acceptable facsimile of the sound of the way I speak. The latter is more

challenging. We had to find ways to 'prompt' or 'train' the LLM to sound more like someone of my (hybrid) demographic: a senior, expatriate Australian, interdisciplinary American academic, speaking in a way such a person would speak in the semi-formal environment of a lecture-room or conference session. That is to say: intellectually rigorous, didactic, seeking to inform a mixed but educated adult audience about a specialized topic, with some interest in keeping them engaged and to some extent, in a rarefied way, entertained. This is clearly a highly specific mode, filtered through a highly specific demographic. Control of 'expressiveness' or emotional modulation became an issue – in some audio generation, a friend observed, I sound like I was on Xanax.

Stage 6.
Hedra deep-fake video tool was used to animated a still-frame of my face (talking head) delivering said vocal tracks, in a video-lecture format. While this was quite successful in many cases, several anomalies occurred. In the source image, my mouth is closed. Hedra 'knows' that faces have teeth that are visible when the mouth opens, so it invents teeth in such cases – every time the mouth opens teeth are a different shape and sometimes, disturbingly, they move about. In some cases, expressiveness became an issue. The face was too 'expressive' and changed expressions rapidly, resulting in a ghastly comical effect.

Stage7.
Slides were generated for the lecture, driving LaTeX from GPT. GPT is a language models, is it not a graphic design tool, so initially the system created generic, graphically simple slides. The problem with this was that they were not like the slides I would make, in terms of typography and graphic design. I have decade of experience in such things, so have developed a particular graphical style. It took some kluging to generate a slide style, that while not looking like slides I might make, was close enough to be 'Simon-like'.

Stage 8.
The least successful aspect of the process was the generation of diagrammatic images to illustrate the slides. This was done by Interfacing GPT with DALL-E. This resulted in the kinds of AI-generated imagery that has become a source of mirth. In some cases truly dystopian images of classroom scene looking like forced-labor camps, in others a proliferation of incomprehensible icons juxtaposed and jumbled together in meaningless ways, all decorated with absurd misspelled captions and usually misformed typography.

Stage 9. GPT 4.o.1 – the genie is out of the bottle
Alarmingly, the newest version of ChatGPT (4-o1) we worked with has shown an increases capability to form extended arguments – previously there were jarring jumps of logic and narrative continuity between paragraphs. It has introduced more sophisticated rhetorical behaviors, positing rhetorical questions and hypotheticals into its style. The newer iterations appears to make a worryingly accurate representation of my academic style. As noted, this has demanded a novel kind of editorial analysis: 'it sounds right but is it actually saying anything (rational, rigorous and novel)?' Still, maybe its too much to ask

ChatGPT to actually have ideas. And that may be a good thing. See https://simonpenny.net/works/AISelfPortrait.html

Stage 10. Multilingual Simsimon
We are currently exploring having Simsimon 'perform' in other languages, the current experiment is Chinese. Pleasingly, Simsimon speaks Chinese as a non-native speaker, with the kinds of pronunciation and grammatical errors one might expect of such a person.

References

Agre, Philip E. Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI. *Bridging the Great Divide: Social Science, Technical Systems, and Cooperative Work*. Geof Bowker, Les Gasser, Leigh Star, and Bill Turner, eds. Erlbaum, 1997.

Audry. S. Art in the Age of Machine Learning. MIT Press 2021. ISBN: 9780262046183

LaFarge, A. (2014). Social Proxies and Real-World Avatars: Impersonation as a Mode of Capitalist Production. *Art Journal*, *73*(4), 64–71. https://doi.org/10.1080/00043249.2014.1036613

LaFarge, *A. Sting in the Tale – Art, Hoax and Provocation*. Doppel House Press, Los Angeles, 2021

Manovich, L. (2022). *AI and myths of creativity. Artificial aesthetics: A critical guide to AI, media and design*. Accessed 31Dec, 2024, https://manovich.net/index.php/projects/ai-myths-of-creativity

Penny. S. *Making Sense -cognition, computing art and embodiment*. MIT Press, 2017.

Penny. S. *Sensorimotor debilities in digital cultures*. AI&Society 2021.

Penny. S. Living in Mapworld: Academia, Symbolic Abstraction, and the Shift to Online Everything https://constructivist.info/18/2. March 2023.

Peraica, A. (2023). Large Datasets and the Particularity of Art: Will There Be Any Art in the Deep Learning Age?. In: Tam, Kk. (eds) Sight as Site in the Digital Age . Digital Culture and Humanities, vol 5. Springer, Singapore. https://doi.org/10.1007/978-981-19-9209-4_7