Demystifying AI with uncertainty in design

by Simone Cibrario Nona   2023

Abstract

The following paper seeks to explore enchanted determinism (Campolo & Crawford, 2020) in machine learning through the lenses of critical technical practice (Dourish et al. , 2004), speculative game design (Coulton et al., A. I. , 2016) and animist design (Marenko & van Allen , 2016), and the critique of AI written in Unlearning Algorithms (Soon et al. , 2020). Two prototypes have been developed with the purpose of creating awareness on the phenomenon of enchanted determinism, sparking critical thinking in users towards inputs and outputs of a machine learning system. This has been achieved with designs that embrace that uncertainty intrinsic to the medium.

Introduction

Just a couple of days prior to me writing this paper, a friend of mine wrote me an innocent message on a social platform to share something cool she found. It was a cute Instagram profile called planet_fantastique[1], filled with pastel colored images evoking nostalgic 70's vibes. The whimsical aesthetic was not the only thing that those images had in common: they also displayed the characteristic uncanny sheen of AI generated images; they looked too perfect to be real, they lacked the grime and texture that our world possesses, and even digital cameras struggle not to capture. This is not however a critique on the quality of those frames, but to how my friend perceived them.

"Profilo interamente generato dall'AI"

Which can be translated to "Profile entirely generated by the AI". She wrote so even though in the very profile description the following can be read: "AI exploration by real human @l.i.z.a.r.d.q.u.e.e.n". Despite the funky name, it was very clear that there was meat and bone behind those aesthetically pleasing pictures. My friend, who is a digitally educated designer, was unconsciously using language that perpetuates a fundamental misunderstanding in how machine learning operates, as noted in The Other Side: Algorithm as Ritual in Artificial Intelligence (Browne, Swift, 2018).
This innocuous anecdote shows how the general public doesn't think of machine learning systems as a Human-Machine interaction (Agüera y Arcas, 2017), but they almost give full autonomy and consideration to the machine, even when it's explicitly stated otherwise. To

---

[1] https://www.instagram.com/planet_fantastique/

quote Pasquinelli: "The expression "art created by AI" actually means that a human operator applies the generative modality of neural networks after training them with a given dataset." (Pasquinelli, 2019) On a global scale, this creates a perception of AI as a mystical subject, veiled in an magical aura of determinism which makes it impervious to criticism: this phenomenon has been baptized by Campolo and Crawford as enchanted determinism (Campolo & Crawford, 2020). They define enchanted determinism as "… a discourse that presents deep learning techniques as magical, outside the scope of present scientific knowledge, yet also deterministic, in that deep learning systems can nonetheless detect patterns that give unprecedented access to people's identities, emotions and social character." (Campolo & Crawford, 2020)

We propose that through design it is possible to shine a spotlight on this enchantment. A critical design piece could embrace the quasi-esoteric nature of the media and provoke thoughts in the users on how its nature is not necessarily deterministic (Soon et al. , 2020). To do so, we developed two accessible browser based prototypes developed using open AI libraries for P5.js and Javascript, both having the explicit purpose of manifesting the enchanted nature of AI technology, creating an association between the vague divination techniques of old and the "thought as deterministic" nature of AI.

Concept Development & Hypothesis

The design process was different for the two developed prototypes, as were the teams involved in development, but both of the projects had the teams face to face with the realities of working with AI in an era of easily accessible machine learning methods and APIs: despite the lack of previous knowledge on the subjects, creating applications revealed to be possible in a great variety of ways, from text to images and all in between. The concepts were brainstormed in a large group of around twenty designers, which then selected the ideas they would most like to explore. This created a scenario where the end result is more of a convergence of similar methodologies applied in the same context than a precise vision for a shared purpose between the designs. However different the circumstance of development may be, the emergent considerations and applications still pointed to the defined objective of challenging the enchantment and mysticism surrounding artificial intelligence.

The subject matter of the two softwares makes it fairly apparent: the first one we developed, MessIA, challenged the user to think about current messianic figures in politics and technology through the use of incoherent data bases. The second one, The Intelligence, let the users play with tarot and divination by employing four different AI methods.

What they share is clearly the interplay between the rational and the irrational, the divine and the computational. They both serve our hypothesis that deep learning methods are heavily enchanted technologies, and carefully constructed critical designs can make this apparent, especially using this self-evident contrast.

How technology can cast a spell on us has been discussed by both Gell, who saw in the enchantment of technology the power of technical processes to bewitch the observer (Gell, 1992), and by the already cited Campolo and Crawford.

AI can also similarly be seen as a charismatic technology, one that holds a "power that goes beyond mere form or function to stimulate devotion, yearning, even fanaticism." (Ames, 2015)

The described fanaticism has already been observed in tech environments (Ames et al. , 2015), and it might actually prevent, without necessarily being a doomsday prophet, "these technologies from becoming part of the messy reality of our lives, rather than helping us transcend it." (Ames, 2015)

An uncritical approach and an enchanted vision of machine learning has other dangers. Pasquinelli cites some of them in his paper How a machine learns and fails (2019): how statistical correlation is often used to explain causation, how biases might occur in multiple steps of the process and how error is rarely considered both in the discourse and in the application (Pasquinelli, 2019). Adversarial images have also been often used as evidence to poke holes in the seemingly impenetrable shield of statistics protecting deep learning, in particular how these small variations can be imperceptible to human eyes (Campolo & Crawford, 2020). The general worry here is that we might observe an application of AI that exerts "power without responsibility". (Campolo & Crawford, 2020)

"Part of the mysticism of neural networks is that they offer answers but not explanations. And for the most part, we don't care—so long as the algorithm performs well on the test set. We consult neural networks like the ancients consulted oracles. The process may be esoteric and opaque, but if it gets results that's good enough." (Browne & Swift, 2018)

Having established the problematic nature of this bewitchment, we would like to spend some words on the subject of uncertainty. To Gell uncertainty is the key element in which magic manifests itself in art and technology (Gell, 1992). A job requires skill, knowledge, work and uncertainty; magic fills that void created by a lack of understanding and control. "The problem of uncertainty is, therefore, not opposed to the notion of knowledge and the pursuit of rational technical solutions to technical problems, but is inherently part of it." (Gell, 1992) Uncertainty is evoked in the following project twofold: on one end we seek to showcase it in the methods, though as absolutely deterministic, on the other end they confront the users with their own impossibility to predict the future. There is an intended design metaphor here that aims at "broadening the cognitive spectrum of the actors involved in the interaction". (Marenko & van Allen , 2016) Marenko and van Allen also proposed how uncertainty can be effectively used as a disruptor in design: "The value of uncertainty and instability advocated here resides in their role in shifting and reconfiguring established perceptions…". (Marenko & van Allen , 2016)

Methods

The main methodology adopted in this research endeavor was the critical technical practice first developed by Agre (Dourish et al. , 2004). However content from other writing was considered and applied. The Games as Speculative Design (Coulton et al., A. I. , 2016) paper written by Paul Coulton, Dan Burnett, Adrian Gradinar was integrated with the critical approach towards artificial intelligence described by Winnie Soon and Geoff Cox in their book Aesthetic Programming: A Handbook of software studies, in particular the chapter on Machine Unlearning (Soon et al. , 2020). Agre created the methodology to highlight the problem of underlying ideologies specifically in relation to AI studies. In the introduction to his paper he writes: "AI has never had much of a reflexive critical practice, any more than any other technical field." (Dourish et al. , 2004) This acknowledgement of AI as worthy of particular social critique makes it particularly relevant for this project and the current socio-technical environment, where it seems that recent development in deep and machine learning only accentuated the enchantment at play. Agre also writes how researchers approaching the subject should have "...one foot planted in the craft work of design and the other foot planted in the reflexive work of critique." (Dourish et al. , 2004) This encapsulates the approach followed during the whole project, deconstructing the technologies examined and tackling them with a critical eye both in the development process and as a design goal. Now more than ever there is a need for a deep critical examination on AI applications, and for designs to dispel them and shine a light on widely spread assumptions on their ontology. To better examine this particular subject, I called upon Machine Unlearning:

"What is learnt should not be separated from the means by which it is transmitted, nor the direction of travel from human to machine or from machine to human. More to the point, the production of meaning lies at the core of our discussion, as are concerns about what is being learnt, and to what extent this has been compromised or inflected by reductive ideas of how the world operates." (Soon et al. , 2020)

This excerpt from the chapter encapsulates the core statements made by the author: when dealing with machine learning it is key to examine what is learned, how it is learned and by whom. Blindly accepting the end results can lead to an epistemological dissonance between the data and the output, both in the eyes of the developer, which might ignore or deny glaring fallacies in the system, and in the end user, who will be further enchanted. The author ends by asking how different uses of machine learning lead to different modes of knowledge production (Soon et al. , 2020), and this is exactly what this project hopes to do: embracing the enchanted nature of this deep learning to dispel its deterministic aura.

Finally this project had an attitude towards time and speculation almost antithetical to that described in *Games as Speculative Design (*Coulton, Burnett, Gradinar, *2016)*. What allows us to draw from this methodology is the shared interest in letting the user consider and in a way experience the realm of futurity. But while the paper puts a huge emphasis on remaining in the realm of the possible when approaching time during a speculative research, our intention was to instill a new sense of time, a sort of enchanted and

prophetic time, which mimics the non-determinist properties of the media involved (Soon et al. , 2020).

Prototype presentation

The exploration of divination through the enchantment of technology led us to develop two prototypes, that in a way built onto each other and tackled similar subjects, while using different applications of machine learning. The first one is MessIA, an app which randomly gives to the user a piece of wisdom generated from an intentionally ambiguous library. The second one, The Intelligence, is an AI driven tarot reader that employs generative methods for both text and images, as well as feeding as input the results of classification algorithms.

MessIA

Users of MessIA will be asked, in a calming minimalist interface reminiscent of silicon valley aesthetics, to press a giant button in order to record their thoughts and feelings. Once by releasing the button the registration is terminated, the user will be presented with an AI generated message. The message will be displayed with slowly unwrapping text and spoken by a robotic text to speech in a deep male voice. This message will take the shape of religious parables, but showcasing an idiosyncratic uncanniness: twitter handles and neo-liberal political statements might blend seamlessly with biblical references. The user can then decide to record a new request to receive a new word of wisdom from the system. The implication here is that the app is listening to the user and then generating real time a custom message based on their input. The reality of what is actually happening is much simpler. The recording function is a dummy feature, whose sole purpose is to trick the user into believing the relevance of their input. The message is also not generated in real time, but fetched from a database of strings. This database has been curated by us from a selection of sentences created with a GPT2[2] based model trained on a purposely conflicting set of data. The curation has been redacted with the goal of filtering the less intelligible messages and highlighting the more outlandish results in terms of mixing the input data in grammatically correct sentences. The original input database included a diverse selection of holy text excerpts from many religions (the Bible, the Quran, buddhist writings etc.), mostly in the form of short verses and parables. The input included, however, also some more contemporary messianic figures: belonging in the world of philosophy and economics, like Marx, male right activists, like Andrew Tate, and prolific and divisive twitter users, like D. Trump and E. Musk. Of the latter two individuals raw tweets have been inserted in the database, without removing any platform specific formatting artifacts such as the twitter handles and hashtags. The prototype ended up working as intended, even though the interface, originally meant for a smartphone, incurs some scaling issues when opened with a browser.

---

[2] https://openai.com/blog/tags/gpt-2/

mess I.A.

fig 1: opening screen of MessIA

The Intelligence

This tarot reading browser application is the culmination of the research conducted in exploring AI in relation to prophecies. The Intelligence uses four different machine learning methods all concurring in reconstructing a divination session based on the popular cards.

Going step by step to deconstruct the inner workings of The Intelligence, the user is asked to snap a picture with their front camera. This can be done both on mobile and on a laptop. The image is used as an input for two different algorithms. First an object recognition method determines what is present in the frame, these labels are saved and stored for later use in the process. The same input is then processed through a Teachable Machine[3] method we trained with over one hundred decks of tarots: this allows us to determine the closest tarot to the taken picture. This ends the phase of data gathering of the application: the program now knows the objects present in the frame and the name of the associated tarot. Subsequently, The Intelligence uses this data to display the image of a tarot, the name of the identified tarot, and a pre-generated message associated with that particular tarot. The image is created using the DeepAI API[4]. The API feeds the AI with a string composed of the labels derived from the object recognition plus the stylistic prompt "in the style of a tarot card".

The final string sent by the API may look similar to this:

"**Person** *and* **Bottle**, *in the style of a tarot card*"

While the image is generated, the program uses the identified tarot to fetch from a text file one of six sentences belonging to that tarot. These verses are not generated in real time, and have been produced with Laika, a GPT based tool for writers. The Laika model has been trained with tarot readings gathered from reddit and other sources. The output has then been curated to filter any eventual errors in the syntax. The final outcome is a unique tarot card, with a recognizable name belonging to the traditional decks, but with a

---

[3] https://teachablemachine.withgoogle.com/
[4] https://deepai.org/

generated image mirroring the picture taken in the content but not in the form. A vague but coherent message will accompany it. The fact that we are not generating the text divinations as the app runs, means that frequent use of The Intelligence will dispel its magic, but that is a constraint we acknowledged during development and we found sufficient in achieving our stated goals.
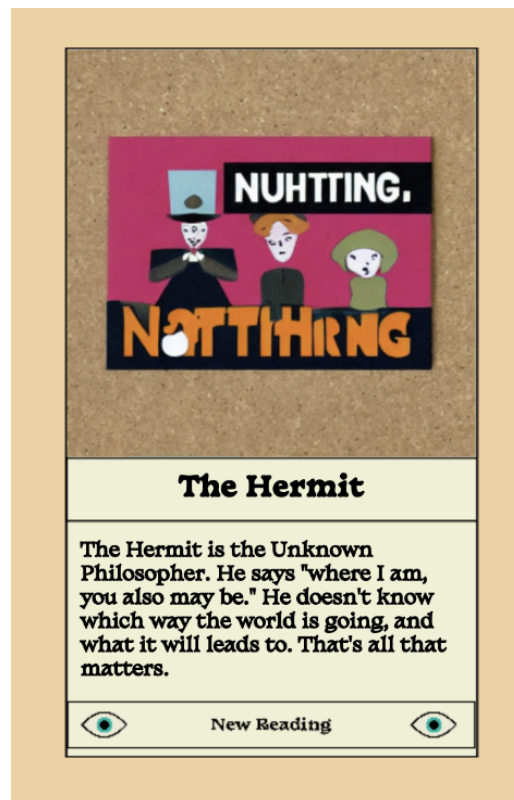

fig. 2 the end result of a session in The Intelligence

Implications for Design

Concerning the user experience, while it was impossible to conduct extensive testing due to time limitations, our personal observation of player behaviors suggests that the prototypes were successful in highlighting the mystification of AI through uncertainty and, sometimes, outright confusion.

In MessIA, no one questioned the authenticity of the record button. The player received a message that in form clearly felt the product of a machine, due to the overall lack of meaning and the stylistic mashups. While many of them correctly identified some of the data used in training the model, they were still under the impression that their recording was clearly used in the process, which just wasn't the case. Even more interesting to observe was the dynamics emerging from the interplay between the object recognition system in The Intelligence and its image generation algorithm. The tagging system often produced inconsistent results, misidentifying objects or only being able to detect "a

person". In the case where the software recognised something close to what was actually shown, but not quite the expected result, the users often engaged and questioned the workings of the machine, trying to reverse engineer its perception and associations. Another example of this at play has been observed in the way they explored what image could be associated with which tarot. Players were quick to realize that the Hermit was the most common scenario, and this was likely caused by the pose taken by a person in front of a laptop. Another card which was organically discovered through play was the sun, which users deduced possible to generate by placing a bright source of light in the top half of the screen.

The described player behavior in the case of MessIA shows enchanted determinism at play: the software offers opaque "answers but not explanations" (Campolo & Crawford, 2020), and in doing so it proves the lack of questioning and understanding present in the subject.

The Intelligence instead testifies how with proper design coated in uncertainty it's possible to disrupt "linear predictability" and "broaden the cognitive spectrum" (Marenko & van Allen , 2016) of users in a way that puts into question, following the principles of Unlearning Algorithms (Soon et al. , 2020), both the input and output of a machine learning system.

Conclusions

Our work has shown the necessity of a critical approach towards machine learning, to dispel the enchantment that has been cast upon us in this period of fast expansion. This call to attention doesn't aim to be a luddite argument, devoid of any hope and interest towards this technology. On the contrary, by accepting and learning to recognize AI technologies and their limitations, it is possible to bring them into our lives in new, more magical, ways.

References

- Coulton, P., Burnett, D., & Gradinar, A. I. (2016). Games as speculative design: allowing players to consider alternate presents and plausible futures.

- Soon, W., & Cox, G. (2020). Aesthetic programming: A handbook of software studies (p. 298). Open Humanities Press.

- Dourish, P., Finlay, J., Sengers, P., & Wright, P. (2004, April). Reflective HCI: Towards a critical technical practice. In CHI'04 extended abstracts on Human factors in computing systems (pp. 1727-1728).

- Gell, A. (1992). The technology of enchantment and the enchantment of technology. Anthropology, art and aesthetics, 1992, 40-63.

- Ames, M. G., Rosner, D. K., & Erickson, I. (2015, February). Worship, faith, and evangelism: Religion as an ideological lens for engineering worlds. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (pp. 69-81).

- Browne, K., & Swift, B. (2018). The other side. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems.*

- Campolo, A., & Crawford, K. (2020). Enchanted determinism: Power without responsibility in artificial intelligence. *Engaging Science, Technology, and Society*, *6*, 1–19.

- Marenko, B., & van Allen, P. (2016). Animistic design: How to Reimagine Digital Interaction between the human and the Nonhuman. *Digital Creativity*, *27*(1), 52–70.

- Agüera y Arcas, B. (2017). Art in the age of Machine Intelligence. *Arts*, *6*(4), 18.

- Ames, M. G. (2015, August). Charismatic technology. In Proceedings of the fifth decennial Aarhus Conference on Critical Alternatives (pp. 109-120)

- Pasquinelli, M. (2019). How a machine learns and fails. *Spheres - Journal for Digital Cultures.*

- Deepai. DeepAI. Retrieved February 19, 2023, from https://deepai.org/

- OpenAI. *GPT-2*. OpenAI. Retrieved February 20, 2023, from https://openai.com/blog/tags/gpt-2/

- Google. *teachable machine.* Google. Retrieved February 18, 2023, from https://teachablemachine.withgoogle.com/

- *Planet fantastique (@planet_fantastique).* Instagram. Retrieved February 21, 2023, from https://www.instagram.com/planet_fantastique/

Projects

- MessIA: https://messia.findelostiempos.net/
  Developed in collaboration with Sebastián Martínez Sánchez and Francesca Tremulo.

- The Intelligence: https://theintelligence.netlify.app/
  Developed in collaboration with Julie Tremblay and Daniel Karlsmose Boel.