

Discipline and prescribe: ethics by design in the AI era

@AnthonyMeasure

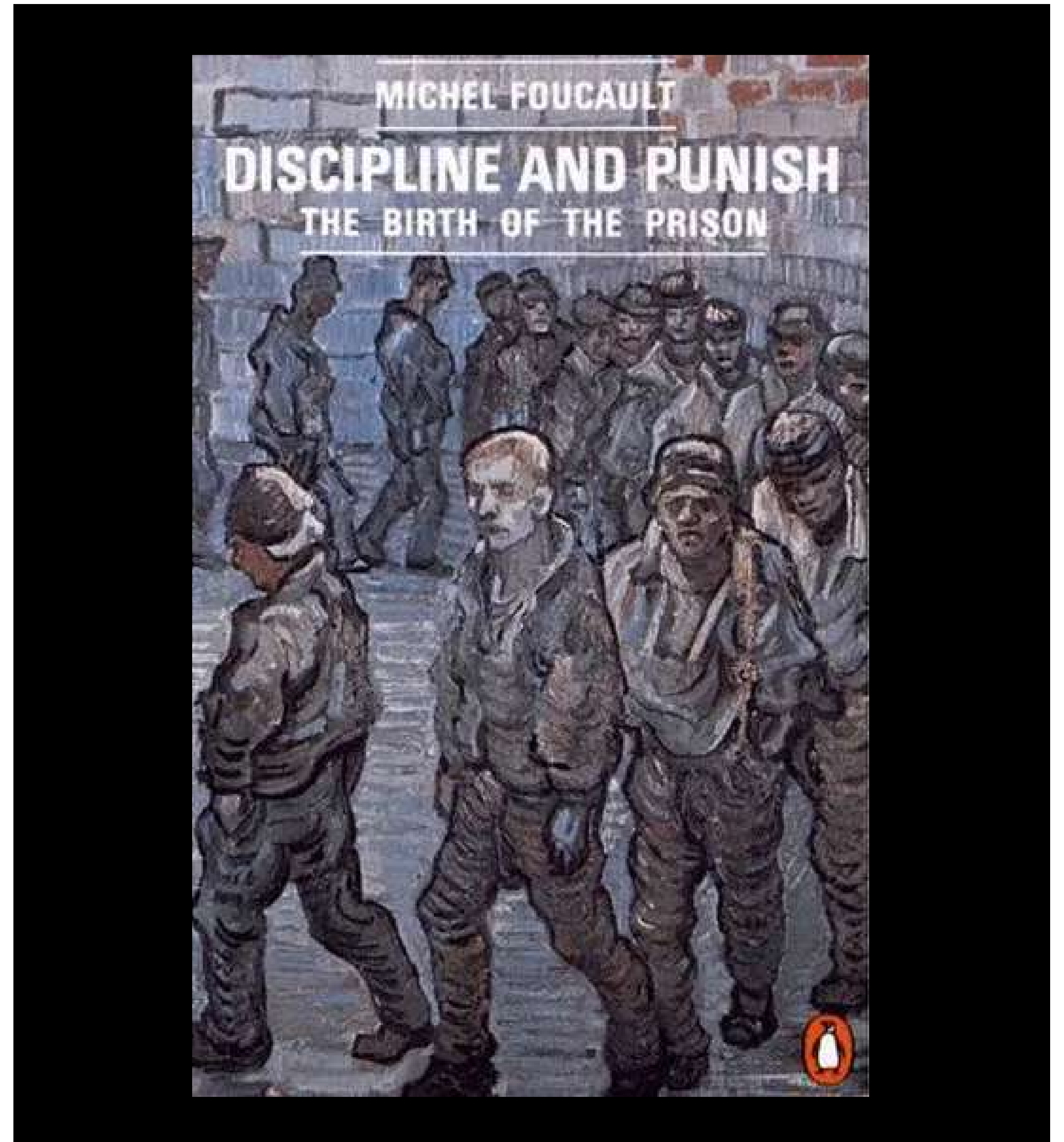
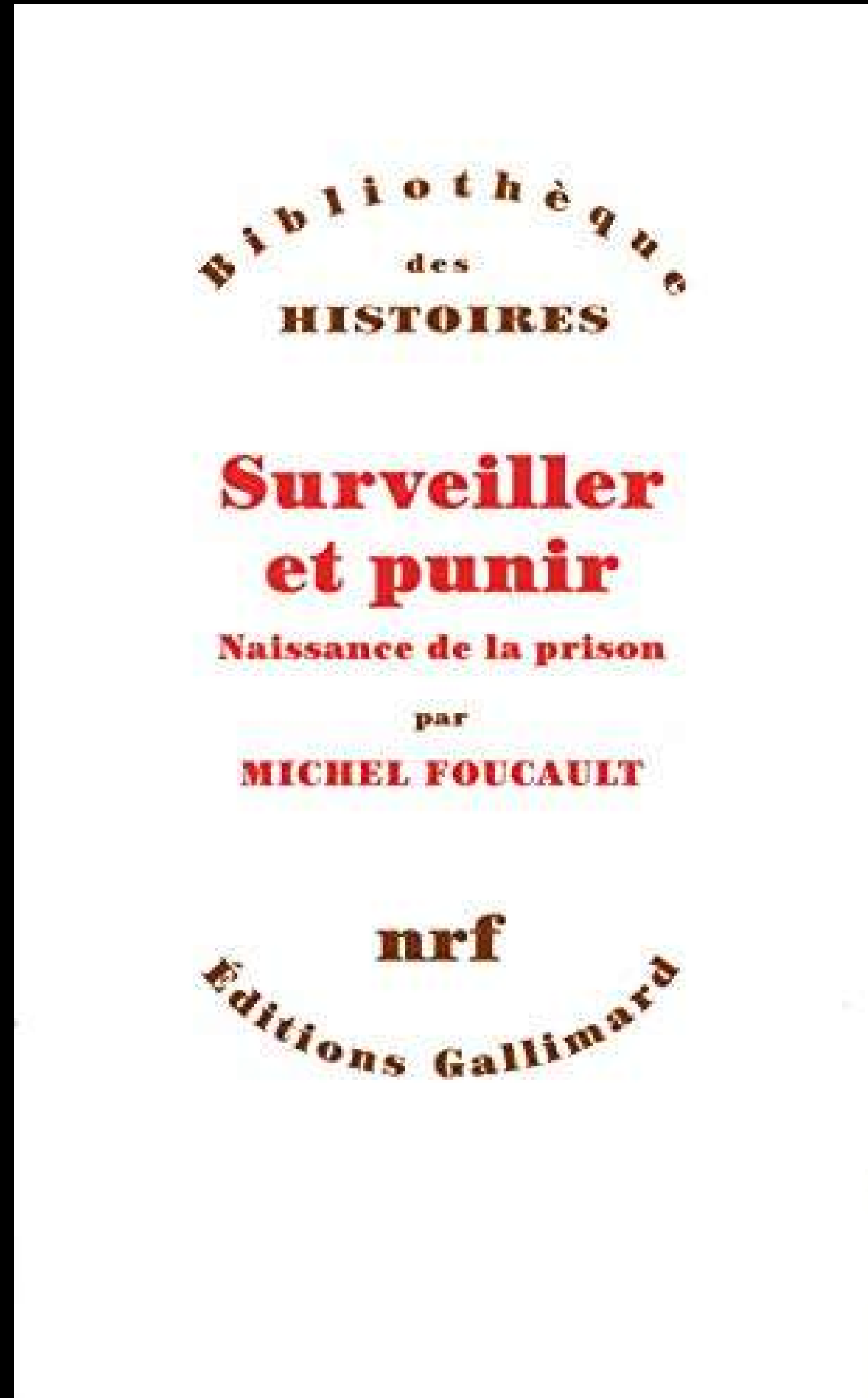
Associate Professor and head of Research, HEAD – Geneva (HES-SO)

—

Lecture for the symposium “Ethics by design in artificial intelligence: theoretical and practical perspectives”

Paris, Unesco EVA Chair on AI, June 8, 2023

Released at the turn of the 2010s, deep learning technologies (a subfamily of machine learning) are based on opaque architectures. As they automate more and more tasks, these black boxes, that are not understandable to humans, overcome the usual ethical and moral categories. Faced with these challenges, some people want to “take back control.” However, if machines (through a mirror effect) reveal the “inhuman” part of humanity, does the human/machine separation still make sense? On the other hand, embedding ethics in a computer program (“by design”) requires making them explicit and fixed—which is the opposite of a dialogical dynamic. To overcome AI challenges, we must therefore “normative ethics” (prescription) and “capability ethics” (potentiation).

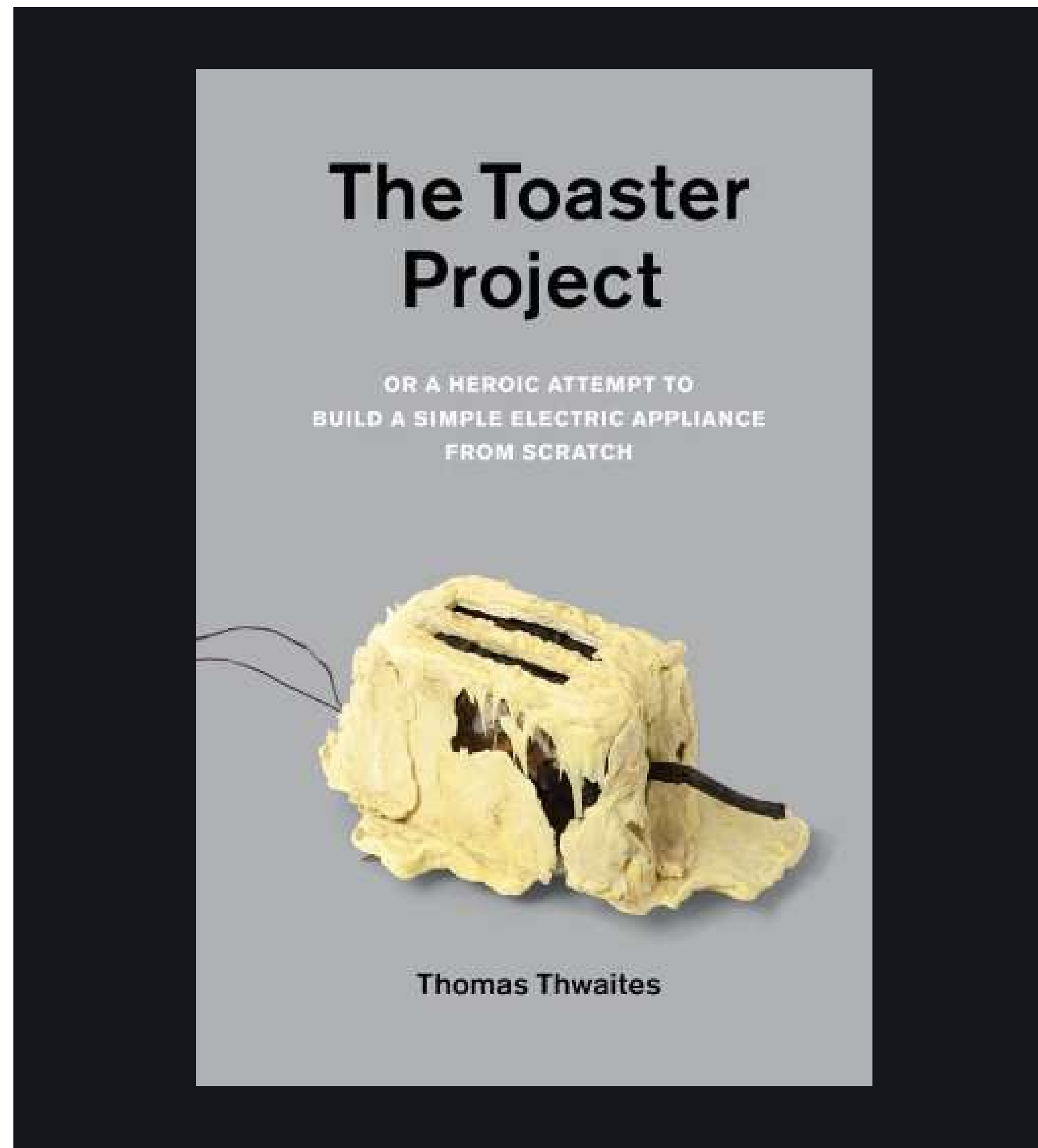


DESIGN?

A field between art & industry

Graphic design, fashion design,
product design, interior Design, etc.

But also: critical design, speculative
design, sustainable design, etc.



1 -

What's the problem with AIs?

Context of the project

"Design & machine learning"

RESEARCH QUESTIONS

- How can AI be reoriented on a smaller scale, in a controlled and "tailor-made" way?
- How can we ensure that AI opens the door to invention and curiosity?
- How could design contribute to defusing the dominant AI culture?



Hes-so
Haute Ecole Spécialisée
de Suisse occidentale
Fachhochschule Westschweiz
University of Applied Sciences and Arts
Western Switzerland

Transition numérique & enjeux sociétaux – Appel à projets exploratoires

I. PROJET – DONNÉES DE BASE

L'évaluation des requêtes se fera de façon anonymisée, c'est-à-dire que les évaluateur-trices ignoreront l'identité des requérant-es. La première page ne leur sera pas transmise. Les autres parties doivent être rédigées de façon que l'identité des requérant-es ne puisse être déduite.

Titre officiel du projet

Design et machine learning : l'automatisation au pouvoir ?

Co-requérant-e (chef-fe de file)			
Nom	Prénom	Statut	
MASURE	Anthony	Prof. associé HES	
Nom de la haute école	Domaine HES-SO	Design & Arts visuels	
HEAD – Genève	tél prof: +33 6 86 98 49 26	Courriel:	anthony.masure@hesge.ch

Mots-clés (max. 10 mots-clés permettant d'identifier des expert-es pour l'évaluation de votre projet)
Automatisation, Création, Deep Learning, Design, IA, Outil, Machine Learning, Programmation

Signature au nom de l'ensemble des co-requérant-es, le/la chef-fe de file :

Anthony Masure, Genève, le 15 juin 2021





Anthony Masure (requérant)

Professeur associé
Responsable Ra&D
HEAD – Genève, HES-SO



Alexia Mathieu

Maître d'enseignement
Responsable du MA Media Design
HEAD – Genève, HES-SO



Douglas Edric Stanley

Chargé de cours
MA Media Design
HEAD – Genève, HES-SO

Research team, HEAD – Genève (2022-2023)

ARTIFICIAL DESIGN

A multisupport and open access essay released by HEAD – Publishing in March 2023.

[www.anthonymasure.com/
essai-design-sous-artifice](http://www.anthonymasure.com/essai-design-sous-artifice)

Formats :
Print, HTML, PDF, ePub, MP3



The opacity of AIs

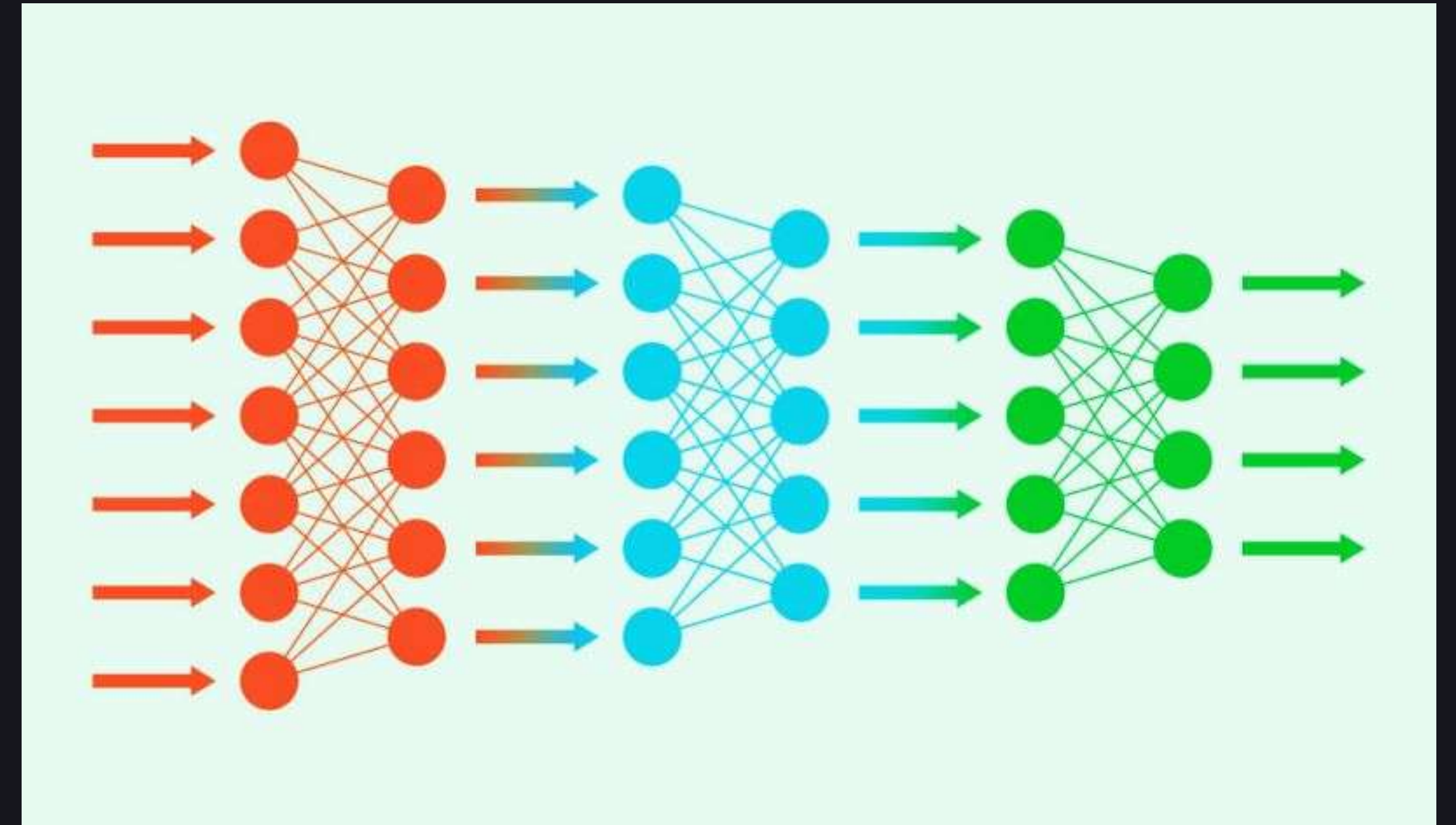
FAMOUS AIS

People have become aware of the development of AI via entertainment and creation-oriented services such as DALL-E (January 2021), Midjourney (July 2022) or ChatGPT (November 2022).



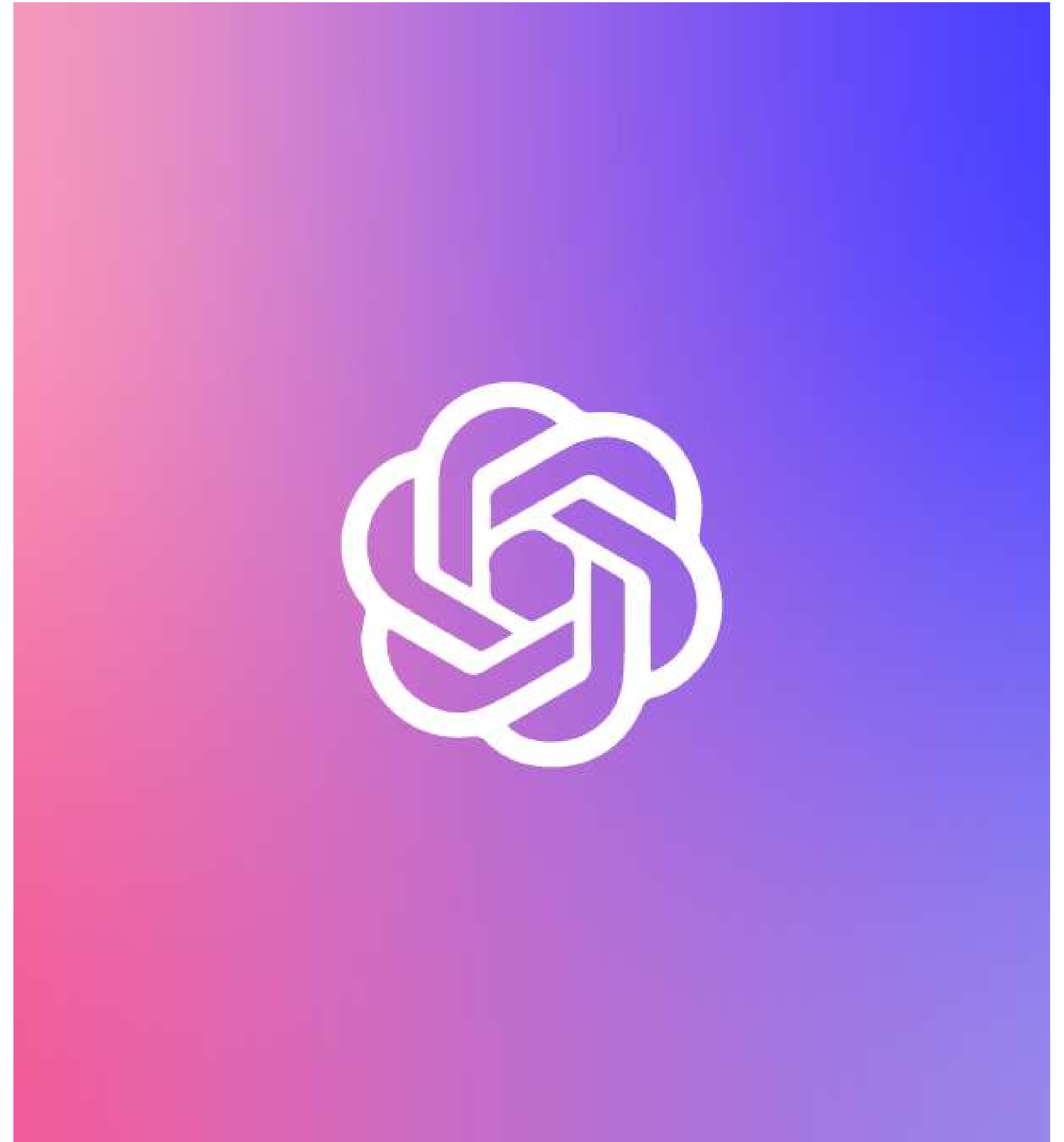
HOW DEEP LEARNING WORK

The machine is given (raw) input data and (labeled) output data. The machine develops a program to match them by “vectorizing” and dividing them into several layers (neural networks).



OPENAI GALAXY (OR OPAQUEAI?)

- Dactyl (2018)
- GPT-3 (2020)
- Codex (2021)
- ChatGPT (2022)
- DALL-E 2 (2022)
- Whisper (2022)
- GPT-4 (2023)

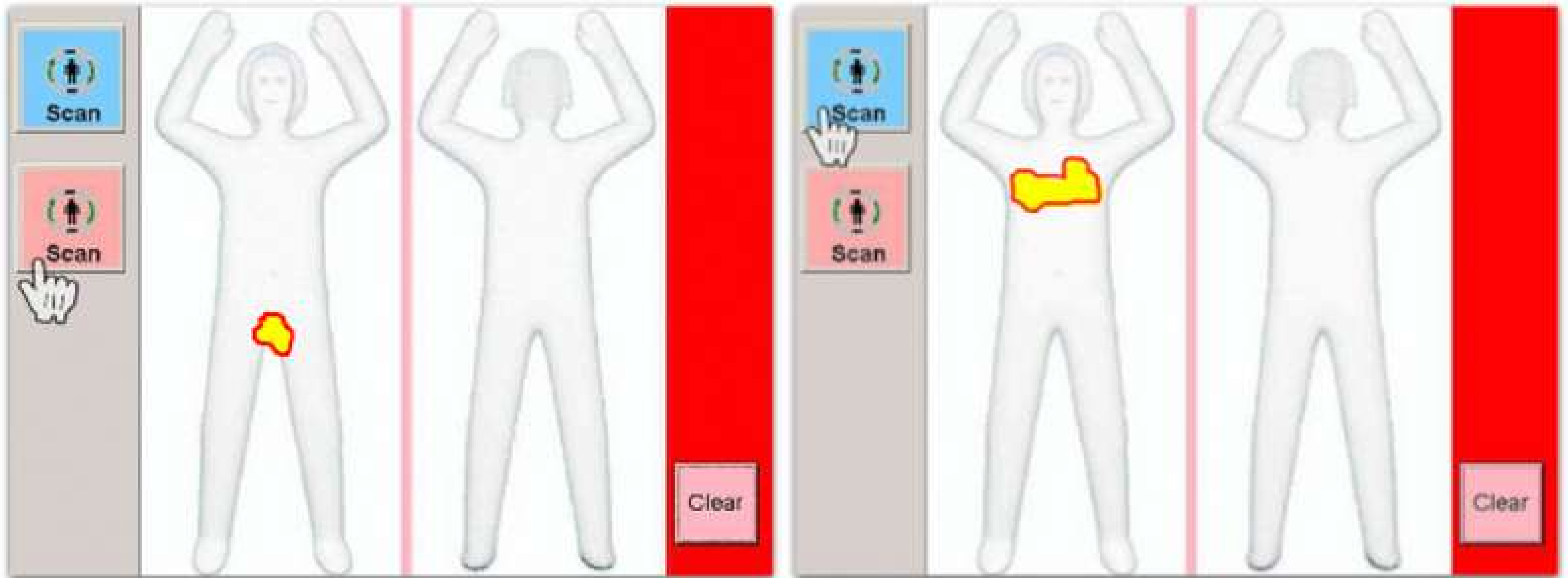




Political issues

“Contemporary AI relates to the ideal of the ‘continuous’ operation of the human psyche, undercutting other approaches, such as psychoanalysis, which, in contrast, are based on the notion of dysfunctionality.”

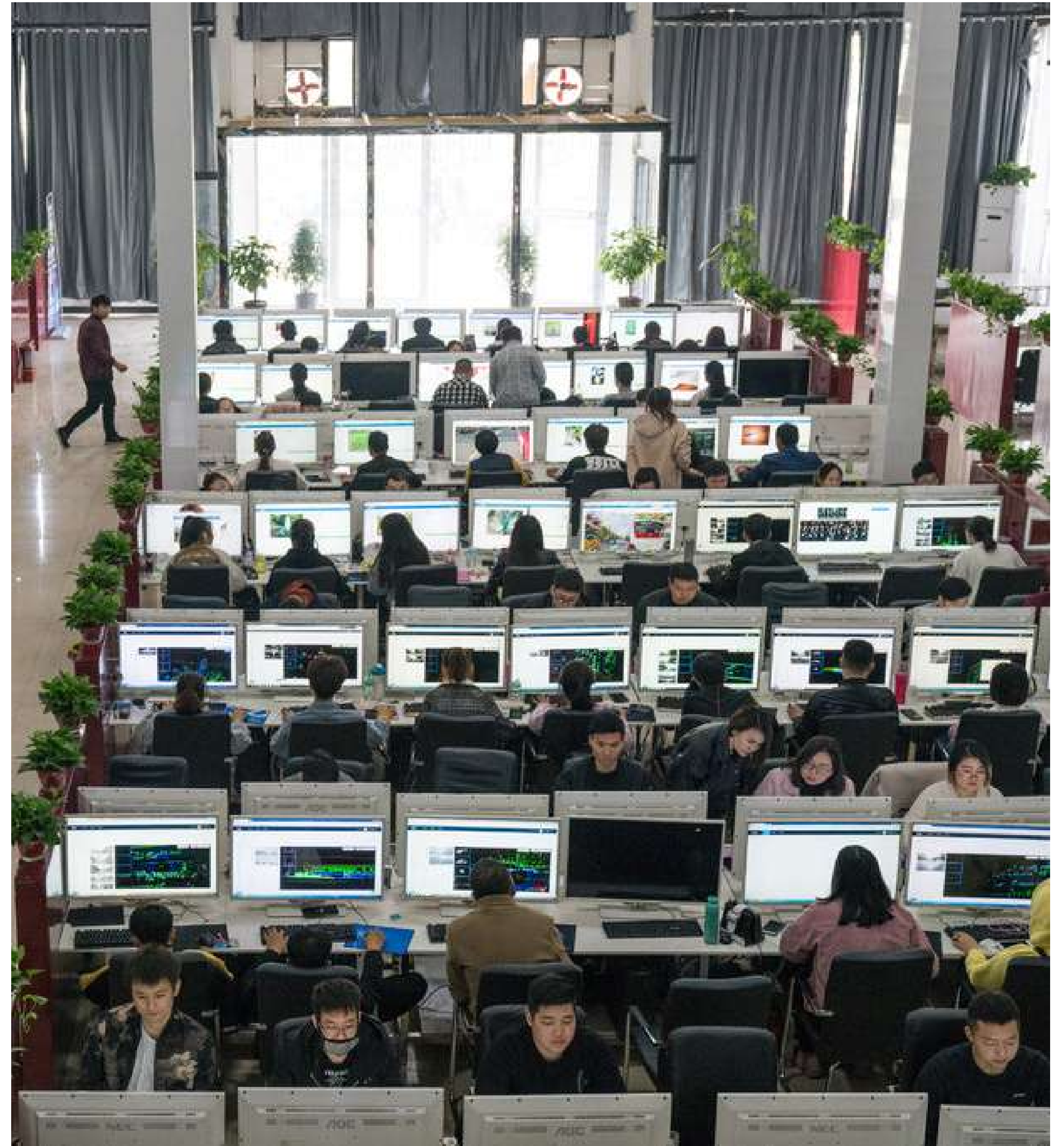
— **Anthony Masure**, *Artificial Design*, 2023



Scanners of trans people at Detroit Airport, 2016

EXPLOITATION OF CLICK WORKERS

Deep learning AIs are not magical programs that work without humans: they require an invisible workforce that sociologist Antonio Casilli calls the “proletarians of the click” (e.g. Kenyan workers improve ChatGPT).



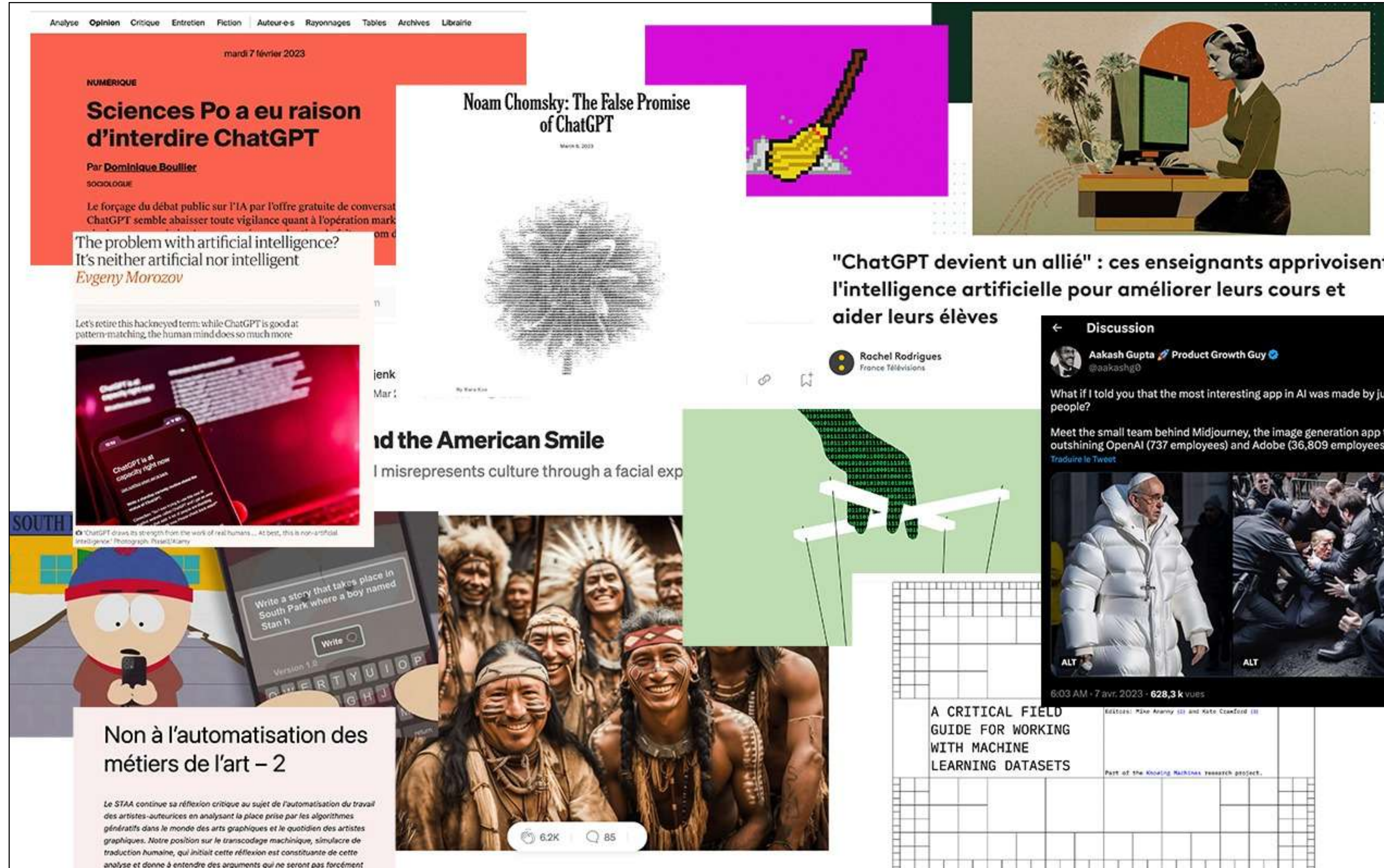
INTELLECTUAL PROPERTY

Programs that automate the creation process retrieve large data sets online without prior authorization. Some artists feel robbed or even dispossessed of their work.





ING, Microsoft, *The Next Rembrandt*, 3D printing, 2016
Revolutionary or reactionary technology?



Anthony Masure, Florie Souday, "State of Art AI & Pedagogy," December 2022 – June 2023
 40 papers: <https://bit.ly/ia-etat-art>

2 -

**Ethics by design to save us
save us from AIs? But what kind
of design are we talking about?**

The dark side of design



“Everything that has transpired has done so according to my design”

Palpatine (alias Dark Sidious) in *Star Wars VI*, 1983

→ Possible negative connotation of the word design in English

Design = machination?

“Everything that has transpired has done so according to my design”
An understanding of design as something that fits perfectly into a plan

“In English, the word design is both a noun and a verb (which tells one a lot about the nature of the English language). As a noun, it means—among other things—‘intention’, ‘plan’, ‘intent’, ‘aim’, ‘**scheme**’, ‘**plot**’, ‘**motif**’, ‘basic structure’, all these (and other meanings) being connected with ‘cunning’ and ‘deception’. As a verb (‘to design’), meanings include ‘to concoct something’, ‘to simulate’, ‘to draft’, ‘to sketch’, ‘to fashion’, ‘to have designs on something.’”

— **Vilém Flusser**, “About the word Design,” trans. from German by Anthony Mathews, London, Reaktion Books, 1999, p. 17

« En anglais, ce mot [*design*] est tout à la fois un substantif et un verbe – détail qui caractérise bien, de façon générale, l’esprit de la langue anglaise. En tant que substantif, il signifie entre autres choses < projet, plan, dessein, intention, objectif >, mais aussi < mauvaise intention, conspiration >, ainsi que < forme, configuration, structure fondamentale >, toutes ces significations et d'autres encore étant liées aux idées de < ruse > et de < perfidie >. **En tant que verbe – *to design* – il veut dire notamment < manigancer, simuler, ébaucher, esquisser, donner forme >, et < procéder de façon stratégique >.** »

– **Vilém Flusser**, « Les bases. Le mot design », in : *Petite philosophie du design*, trans. from German by Claude Maillard, Paris, Circé, 2002, p. 7

« Mais qui donc alors devrait être rendu responsable du fait qu'un robot en vient à tuer ? Le constructeur du robot, du couteau, ou bien celui qui a programmé le robot ? Ne serait-il pas possible aussi d'imputer la responsabilité morale à une erreur de conception, de programmation ou de fabrication ? Et qu'en serait-il si cette responsabilité était mise à la charge du secteur industriel qui produit les robots ? Ou bien encore du [...] système dont ce complexe fait partie ? En d'autres termes, **si les concepteurs, les designers, ne débattent pas de ces questions, il peut en résulter une irresponsabilité totale.** »

— **Vilém Flusser**, “L'industrie du design renferme-t-elle une éthique ?” [1991], in : *Petite philosophie du design*, 2002, p. 26-27.

“Whom should one hold responsible for a robot killing somebody? The person who constructed the robot, the one who made the knife or the one who set up the computer program? Would it not equally be possible to attribute moral responsibility to an error in construction, programming or production? And what about attributing moral responsibility to the branch of industry that produced the robot? Or perhaps the whole industrial complex [...]? In other words: **A situation in which designers do not address themselves to these questions can lead to a total lack of responsibility.**”

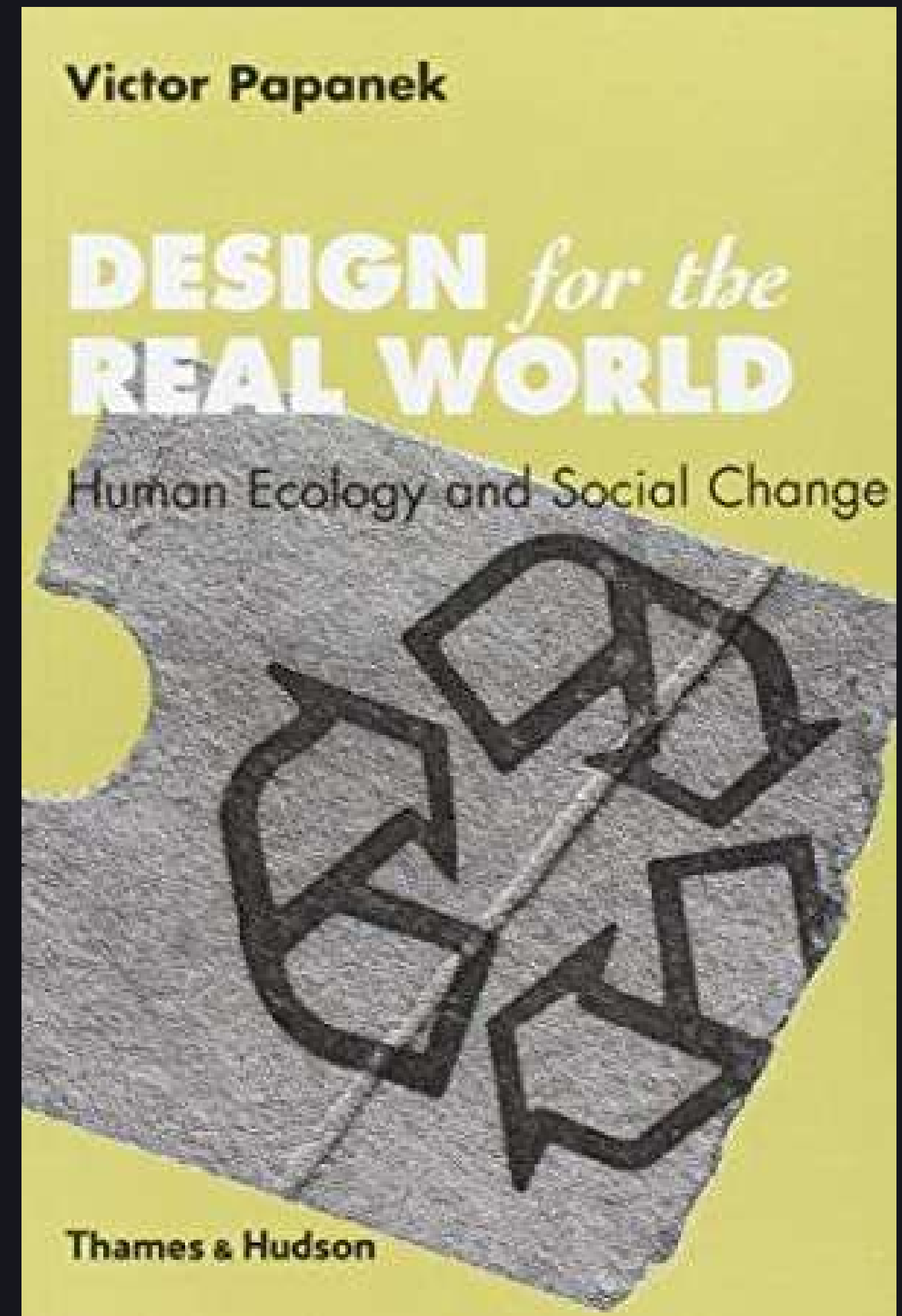
— **Vilém Flusser**, “The lever strikes back,” trans. from German by Anthony Mathews, London, Reaktion Books, 1999, p. 17

« But the question about ethics in industrial design, about the moral and political responsibility of the designer, has become valid (and even urgent) in the present situation. »

— **Vilém Flusser**, « The Lever Strikes Back » [« Le levier contre-attaque »], conference at the Design Academy Eindhoven on April 20, 1991, published in the conference proceedings under the title “Ethics in industrial design? ”.

—

**The history of design is
made of moral crises**



Victor Papanek, *Design for a real world. Human Ecology and Social Change* [1971],
Ed. Alison J. Clarke and Emanuele Quinz, trans. of English (United States) by Robert Louit and Nelly Josset, Dijon, Presses du réel, 2021

“[The book] deals with design for the elderly, disabled, poor, blind, deaf, fat, left-handed, and pregnant; toys and school environments for children; medical instruments and hospital equipment; Third World needs; in short, the book deals with everything that industry and designers currently serving industry have conveniently forgotten. **It deals with the political, social, ecological and environmental effects of design, and the responsibilities that this entails for design teams.**”

— **Victor Papanek**, “Reply to [Guy] Bonsiepe’s Book Review,” 1974

3 -

What “responsibility” in a world where criteria and criticism are in crisis?

**We have to deal with
prescription machines**

Critique, Critères, Crise

Vilém Flusser

Le thème

Une hypothèse sera proposée, selon laquelle notre faculté critique est à présent en crise par manque de critères qui soient adéquats à notre situation culturelle¹. À la racine des termes « critique », « crise » et « critères » que je viens d'utiliser se trouve le verbe grec *krinein*. Ce verbe signifie à peu près : diviser pour pouvoir juger. Il s'agit, dans ce verbe, d'un geste qui casse une unité pour voir ses composants. C'est un geste destructif. C'est pourquoi nous retrouvons le même verbe à la racine du terme « crime ». Et c'est ce geste qui sera le thème de cette intervention.

¹ Ce texte a été rédigé dans le cadre d'une conférence donnée à l'Université de Liège (Belgique) le 29 octobre 1984, à l'invitation d'Hélène Parizel. La correspondance de l'auteur indique qu'il avait, à l'origine, préparé ce thème pour une intervention à l'université de Bielefeld (Allemagne) le 2 novembre de la même année. Ce document porte le n° 3039 (V.2) dans les Archives Flusser; il a été retranscrit du tapuscrit et édité par Gabrielle Contout.

La critique

Pour pouvoir critiquer, il faut prendre ses distances par rapport à la chose. Ses distances, cette « transcendance », n'est pas aisément accessible. Il faut conquérir la distance critique. Cette conquête est un fait historique. La faculté critique a été acquise à un moment donné et à un endroit donné. On peut reconstituer cet événement à l'aide de l'archéologie.

Il y a, dans un dépôt à Ougarit, une quantité de jarres à l'huile. Ces jarres constituent un ensemble, une unité. On peut imaginer cet ensemble : il a une forme, une *Gestalt*. Et on peut garder cette image dans sa mémoire. Il s'agit maintenant d'exporter cent jarres vers la Crète. Pour pouvoir le faire, il faut compter les jarres. Les arracher, l'une après l'autre, du contexte de l'ensemble. Ce n'est pas la peine qu'on le fasse physiquement. On peut le faire symboliquement. On peut prendre l'image de l'ensemble qu'on a dans sa mémoire, et arracher les symboles des jarres (les pictogrammes), l'un après l'autre, du contexte de l'image. Et ensuite,

212

© Association Multitudes | Téléchargé le 08/06/2023 sur www.cairn.info (IP: 46.193.109.226)

Programme (Tes père et mère honoreras)

Vilém Flusser

Ce terme, et son équivalent latin *prescription*, signifient « texte dont le propos est de provoquer un comportement spécifique de la part de son récepteur¹ ». Exemples : les Dix commandements, le Code civil, tel ou tel code de circulation, n'importe quel mode d'emploi, n'importe quel programme d'ordinateur. Cette série d'exemples peut être lue comme un abrégé de l'histoire occidentale. La prescription se désacralise progressivement, pour devenir de plus en plus profane. Le comportement se dépolitise progressivement, pour devenir de plus en plus fonctionnel. Mais, dans cette série progressive, il y a rupture. Toutes les prescriptions, des Dix commandements jusqu'au mode d'emploi, s'orientent vers l'homme. Tout programme

¹ Ce texte constitue la deuxième partie d'un ensemble de deux écrits publiés dans la revue *Théâtre/Public*, n° 67, janvier-février 1986, p. 79-81, sous le titre général « Tes père et mère honoreras ». La première partie, intitulée « Y a-t-il un futur pour l'écriture? », était l'esquisse de l'introduction de l'ouvrage éponyme (encore inédit en langue française). Ce document porte le n° 699 dans les Archives Flusser; il a été retranscrit du tapuscrit et édité par Yves Citton.

190

© Association Multitudes | Téléchargé le 08/06/2023 sur www.cairn.info (IP: 46.193.109.226)

d'ordinateur, lui, indique une machine. Je proposerai dans cet essai l'idée que la rupture entre l'avant-dernier et le dernier exemple de la série est ce qui nous caractérise.

★


C'est dans les plus anciens textes que se trouvent les prescriptions. Elles programment, à l'origine, le comportement humain face à la divinité (les Dix commandements). Par la suite, le comportement humain face à la société (le Code civil). Plus tard, le comportement humain face à la machine (mode d'emploi). La rupture intervenue, elles programment le comportement de la machine face à une autre machine. Ni *Deus ex machina*, ni *Ex machina Deus*, mais : *Machina ex Deo*.

Perspective utopique : dorénavant les programmes seront prescrits aux machines, et les hommes, eux, seront libres. Utopie rationaliste. Plus une machine s'automatise, plus son mode

Vilém Flusser, “Critique, Critères, Crise,” *Multitudes*, no. 74, 2019, pp. 212-217
Vilém Flusser, “Programme (Tes père et mère honoreras),” *Multitudes*, no. 74, 2019, p. 190-193

4 -

**Is design responsible for
the current situation?**

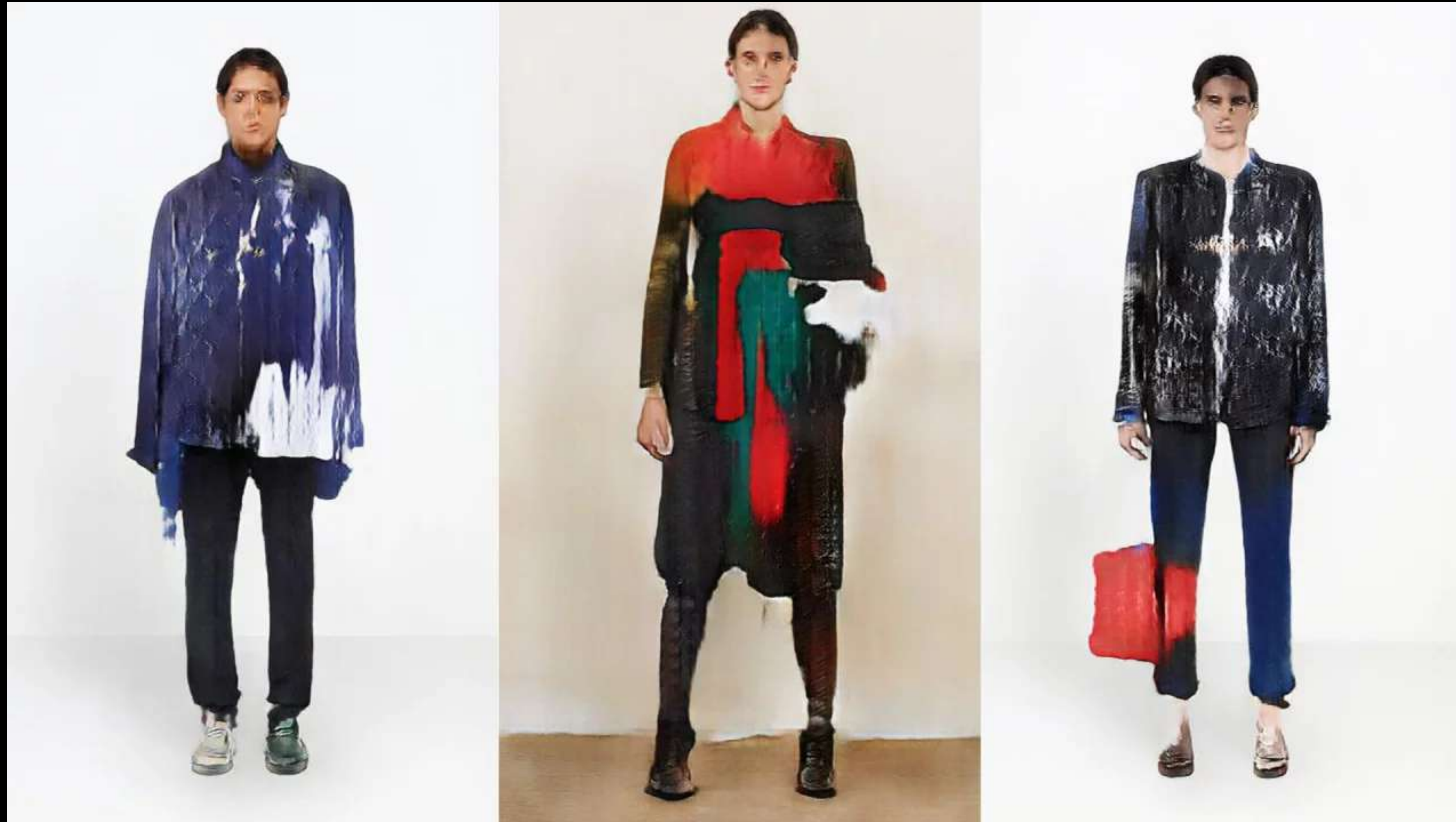


Design as solution or problem?

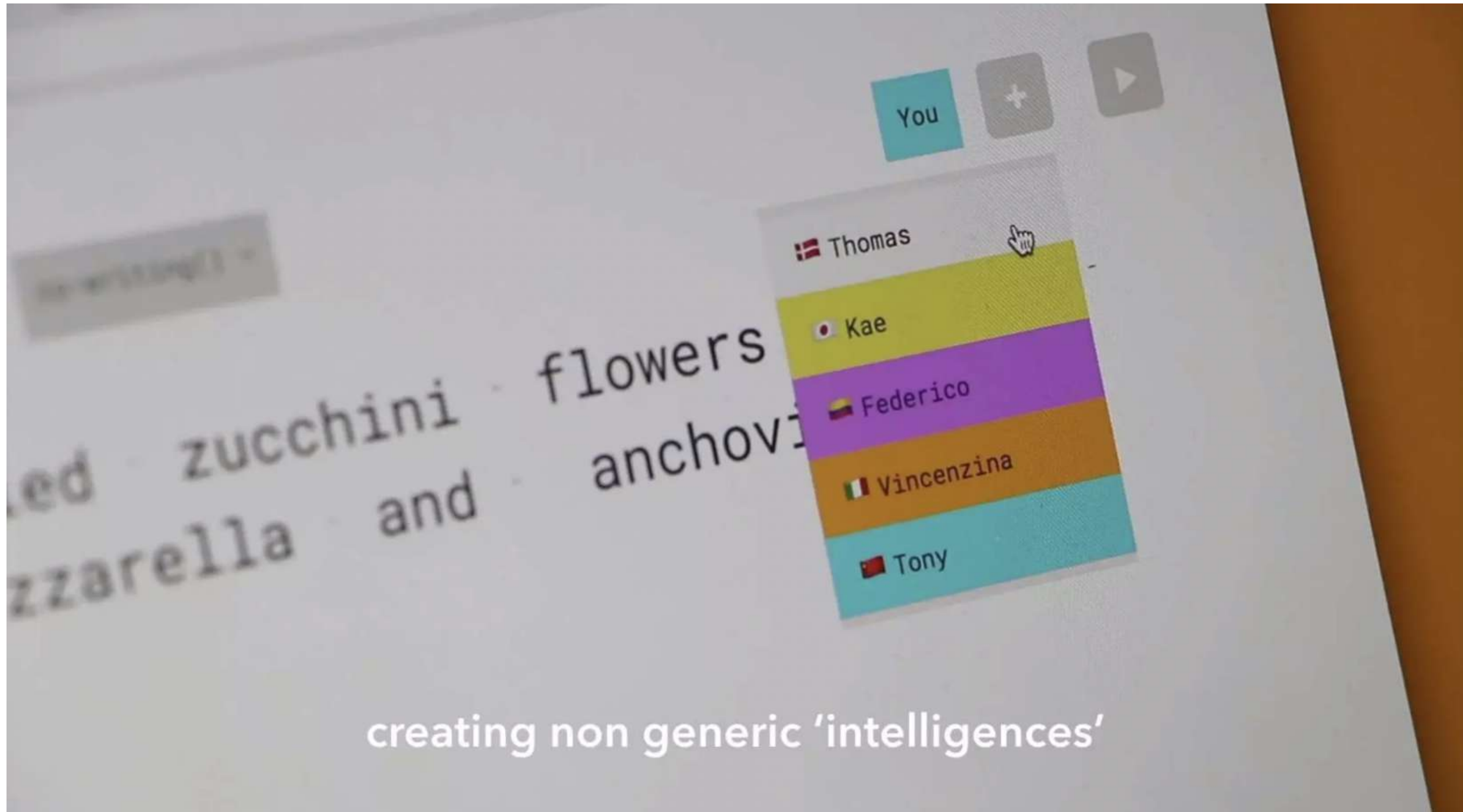
“This [Italian] model refers to **a design that does not correspond to an industrial function limited to solving production problems, but to a strategic activity, to a civil culture**, involved in changing history and therefore able, through the project, to provide large-scale industry with an identity within society. Prophetically, Ettore [Sottsass] sees an open industry that does not seek to integrate society or culture, but which, on the contrary, tends to insert itself into their meshes in order to be a permanent engine of development.”

– **Andrea Branzi**, “La qualité formelle du monde,” in: Collectif, *Ettore Sottsass*, Paris, Centre Georges Pompidou, 1994, p. 27

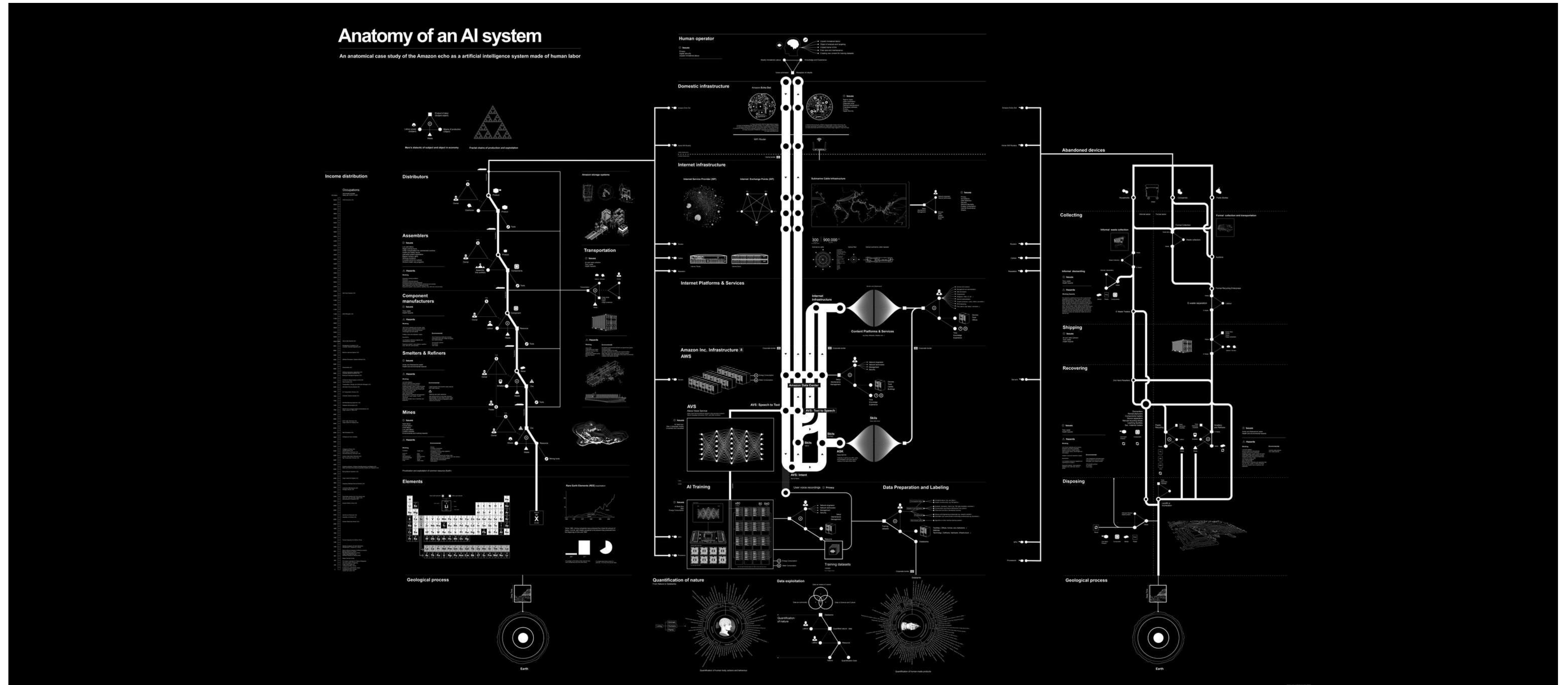
— Counter the dominant AI culture through design



Robbie Barrat, *Balenciaga AI*, 2018



Simone Rebaudengo, *Domesticating Intelligence*, 2016

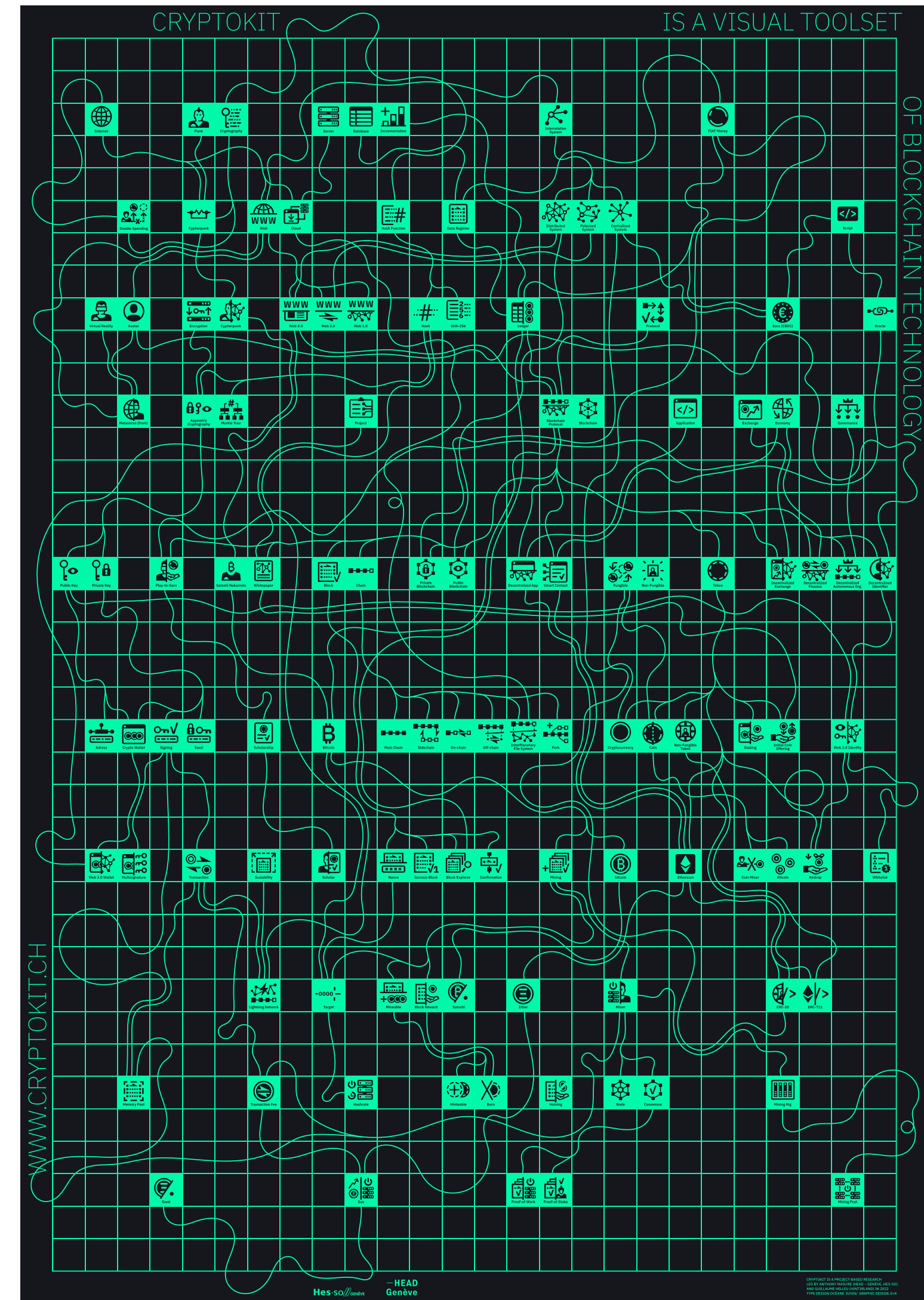


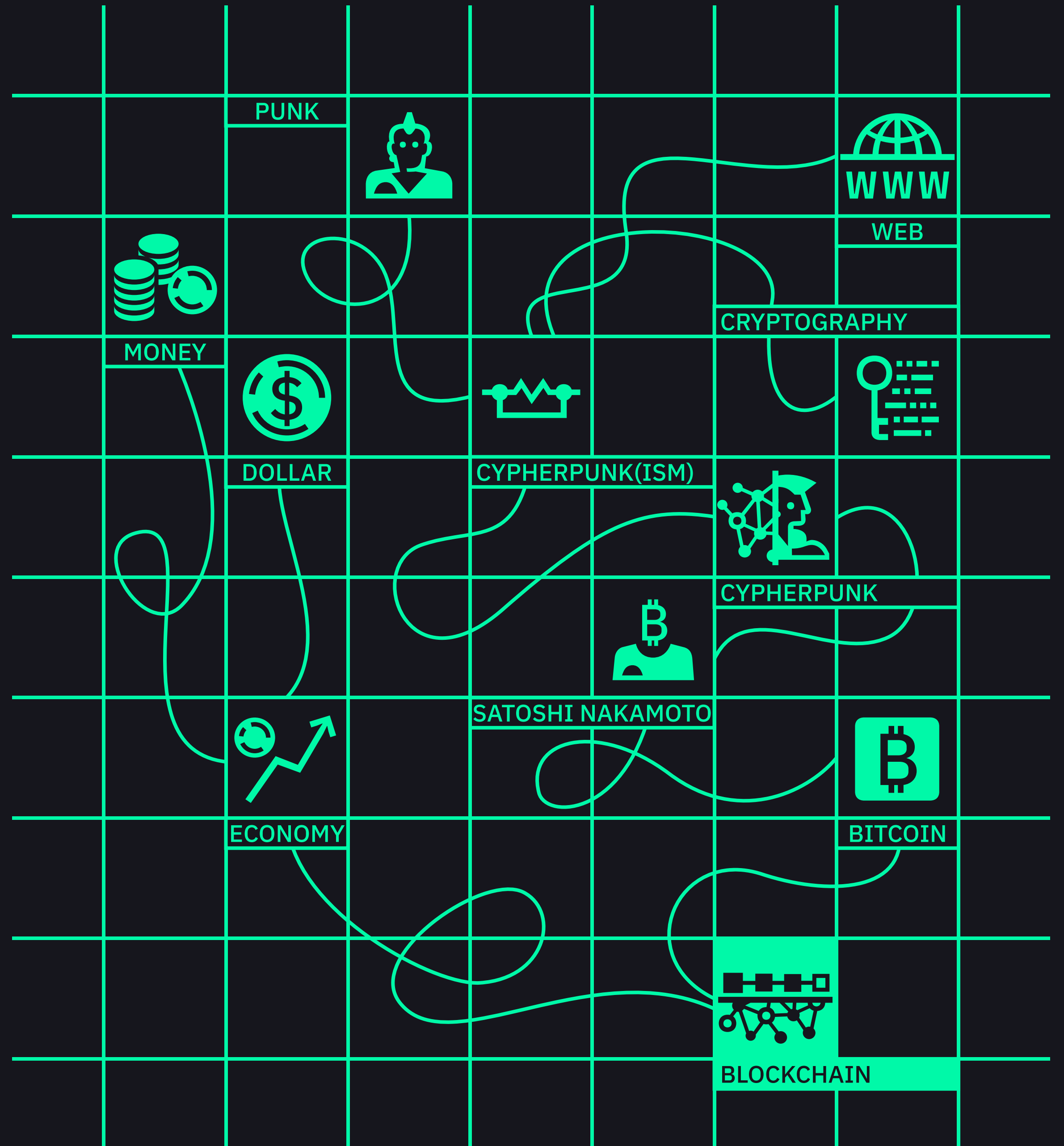
Kate Crawford and Vladan Joler, « Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources », AI Now Institute and Share Lab, 2018, <https://anatomyof.ai>

CRYPTOKIT (2023)

CryptoKit is an open source semantic map composed of 200 pictograms designed to explain blockchain technologies.

Research project conducted at HEAD – Genève in 2022-2023 and funded by HES-SO.







Conclusion

Towards responsibility design

Ethics of conviction and ethics of responsibility

The “ethics of conviction” is based on the Kantian principle of duty: we must act according to higher principles in which we believe.

The “ethics of responsibility” is based on consequentialist philosophy: we must act according to the concrete effects that we can reasonably foresee.

“The dilemma is as follows: on the one hand, justification by good intention alone means removing the secondary effects from the sphere of responsibility, as long as we choose to ignore them [...] . On the other hand, taking responsibility for all consequences, including those most contrary to the initial intention, results in the human agent being indiscriminately responsible for everything, **in other words, responsible for nothing he or she can take responsibility for.**”

— **Paul Ricœur**, “Le concept de responsabilité, essai d’analyse sémantique,” in : *Le Juste*, Paris, Esprit, 1994, p. 66

AI ethics by design?

Embedding ethics in a computer program (“by design”) presupposes the ability to make them explicit and fixed, which runs counter to a dialogical dynamic.

We think that current technological mutations call for **a combination of “normative ethics” (prescription) and “capability ethics” (potentiation).**

Issues of designing responsibility

- In an increasingly complex and unintelligible world (due in part to digital layers), how can we maintain a sense of the shared world and enable informed decision-making?
- How can we balance the reduction of citizens to consumers and/or users?

Fields

- **Legal design:** regulations, litigation
- **Banking:** investments, governance
- **Representative democracy:** political programs, elections
- **Complex technical systems:** machine learning, blockchain
- **Ecology:** carbon emissions, attention ecology

Designing responsibility: ways forward

- **Anthology of texts** on the notion of responsibility in the field of design
- **Case studies of design projects** with different layers of responsibility (legal, social, moral, political)
- **Visual mapping** of ethical mutations and shifting value chains

@AnthonyMasure

[www.anthonymasure.com/
essai-design-sous-artifice](http://www.anthonymasure.com/essai-design-sous-artifice)

www.design-machine-learning.ch

