



“If we only follow the optimized recommendations of logical machines and optimized AI systems, humanity will lose its significance since we will no longer have a framework for mistakes.”

ISABELLA HERMANN

The Manifesto

In 2040, artificial intelligence will no longer be used to optimize human life, but to de-optimize it. This story is about making machines more human instead of making humans more like machines. It is about letting coincidence come back into our lives.

INTRODUCTION

The short story “The Manifesto” shows us how far artificial intelligence (AI) has progressed by the year 2040 from the perspective of an entry from the official “European Political Information Service.” AI basically means optimization. Accordingly, applying AI-based systems when it comes to social life means trying to make our lives more efficient. The underlying rationale of all kinds of social algorithms is that we can solve

social problems technologically if we only have enough data and computing power. But what is the goal of this optimization and strive for efficiency? Is this really what is important for us in our lives, or is it rather the goals of companies and states that are hardwired into the software? Will we become increasingly unfree, other-directed, and intellectually bored without even noticing it? Will humans ultimately become

like machines? In the current reality of 2020, such discussions around AI are taking place within the European Union (EU). The EU wants to assert itself as an avant-garde that tries to establish a third way of ethical “AI made in Europe” in contrast to the “market-capitalist” system of the US and the “state-capitalist” system of China. Despite all the severe political issues affecting Europe and the world right now, just imagine

what would happen if the EU achieved that objective—and achieved it in an original and cheerful way. In this utopian setting, new technology is neither rejected nor uncritically embraced. Rather, machines support our humanness. Is this too good to be true? Maybe, but at least, the European Political Information Service tells us a positive narrative of the future ...

THE EUROPEAN POLITICAL INFORMATION SERVICE

History/European history/legislation/embedded mismatch

Last updated: February 8, 2020

All European citizens have the right to de-optimization. Only a few years ago, this was almost unthinkable. This dossier traces the history of this process and looks at the development of the legislation on the embedded mismatch.

INTRODUCTION

Since *»artificial intelligence (AI)* was coined as a term within computer science in the 1950s, AI hype cycles have alternated with AI winters, i.e., periods of great expectations for technology and periods of reduced funding and research interest. The *»AI-hype cycle of the 2010s*, which was enabled by new technological advancements through *»deep neural networks* and the

availability of *»big data*, was superseded by the *»AI winter of the 2020s*, after it had become clear that these systems were not much better at solving real world problems than their predecessors. The game for AI changed again after a major breakthrough in the field of *»quantum computing* by the *»International Research Center for Quantum Mechanics (IRCOM)*.

After quantum computers became the norm at the beginning of the 2030s, a new AI hype cycle began, leading to a wave of computer systems with unprecedented efficiency. This enabled a pervasive optimization of human life in all spheres—political, economic, or societal. With the help of AI the *»UN Sustainable Development Goals*, which were believed to had been lost in the 2020s, were achieved by the mid-2030s. Finally, global challenges like poverty, climate change, or environmental degradation were brought under control through the use of technology.

However, despite all these immense accomplishments, the difficulties we had come to know in the last AI hype cycle of the 2010s returned in a new form. Governments and businesses had to cope with negative effects, such as algorithm overlearning and the deterministic predictions of AI systems in conjunction with extensive reliance on computer programs. But this time, the question was not how to avoid discrimination resulting from biased data but how to avoid general societal boredom and saturation through over-optimization by the logic of machines. The major problems for human society associated with this new rule of logic were, firstly, the loss of chance and coincidence, and, secondly, the loss of basic problem-solving competencies.

THE AI MANIFESTO

In 2034, these concerns led to the formation of a European group of technological and intellectual forerunners who called themselves the *»avant-garde34*. They came up with a declaration containing a precise description of the problem and a possible solution. They called it the *»AI Manifesto*.

»Source Text«

The AI Manifesto

“Logic, understood as a reasonable conclusion, is a theoretical and normative idea. It is theoretical and normative, because people

don't do it. The way we think is that we make logical mistakes. But our strength lies precisely in the fact that humans are not logical machines: By deviating emotionally and cognitively from logical baselines, we create relevance criteria. And "logical behavior" is a relevance criterion in itself. Relevance criteria give meaning to our actions. Mistakes make us free, because mistakes mean that we can act otherwise and differ from the norm. Deviation from the norm is a manifestation of our freedom. If we only follow the optimized recommendations of logical machines and optimized AI systems, humanity will lose its significance since we will no longer have a framework for mistakes. Life is about making mistakes not avoiding them; without them, life will become unfree and insignificant. However, like mistakes, technological progress is human, too. Therefore, we need nonlogical conclusions to be obligatorily embedded into artificial intelligence systems. Nonlogical conclusions will necessarily inspire discord and unforeseen predictions; they are a tool against optimization that can enrich our lives with coincidence. We call this tool for de-optimization the embedded mismatch."

Many of the members of the avant-garde³⁴ sat in technology committees, task forces, and councils and thus could exert political influence. They strategically used their leverage to promote a regulatory framework to realize the AI manifesto and make the vision of de-optimization European law. The AI manifesto went viral and sparked a comprehensive, substantial, and heated public debate. At the beginning, the general opinion coming from diverse stakeholder groups was negative and dismissive, if not hostile. They found the idea of programming a computer to make mistakes ridiculous, dangerous, and unworldly. For different reasons the joint perspective was that if the embedded mismatch became EU law, it would be the last law before the downfall of Europe.

»Source Text«

Official Stakeholder Quotations Concerning the AI Manifesto

"Humans naturally strive for faultlessness: Technology has always aimed to free people from their errors, evolutionary struggles,

and predatory heritage around competition and status—we don't want to have that again”—»President, European Association for Coders

“Machines still have more to do with artificial stupidity than artificial intelligence. Even with quantum computing, machines only analyze correlations and know nothing about social and cultural contexts. What could happen if we implement embedded mismatches? It might turn out to be a disaster. The use of automated systems should not be promoted by programming more errors; instead, it should be limited.”—»Chair, Algorithmic Transparency International

“We have finally reached an international balance, with open trade and exchange between the great power blocks of China, ASEAN, the US, the African Union, Russia, and Europe. Any sort of de-optimization would be a competitive drawback for Europe. None of the other players would take us seriously anymore; we would make a fool of ourselves and the others would turn their back on us.”—»CEO, Global European Unicorn Investment

“The implementation of something like an embedded mismatch would open the door to all sorts of international espionage and cyber-attacks. Hostile actors might hack systemic infrastructures under the guise of embedded mismatches and gain control over government institutions. The consequences would be severe security and societal risks”—»Head, European Intelligence Agency

“The introduction of illogical conclusions would violate Kant's categorical imperative in new and yet unseen terms, because with this system, people would be treated only as a means and not as an end. Forcing intelligent human beings to follow the stupidities of a nonrational system is against everything the Enlightenment has brought and taught us.”—»Director General, Digital Kantian Ethics Department, University of Europe

While the major stakeholder groups from the business, research, civil society, and security sectors refused the AI Manifesto, the two great political movements in Europe—the »*populists* and the »*elitists* embraced the idea as a way to ensure human flourishing, given that artificial intelligence would likely be persuasive in modern societies. The populists agreed that the embedded mismatch could be an opportunity to limit the power of an elite-inclined mainstream, whereas the elitists—being concerned that

a bored and saturated people would be a danger to democracy—saw the proposed de-optimization as a way to ensure freedom and political stability. Key players from the European Government and the European Parliament came to believe that humans had been creating ever more powerful technology without a genuine and benign human-centered philosophy for long enough. There was agreement that de-optimization by the embedded mismatch was not directed against Enlightenment but was a means to realize the »Enlightenment of the twenty-first century as the new philosophical guide. It became political common sense that the Enlightenment of the twenty-first century could not be a matter for ethics alone but needed to be addressed by rules that were enforceable and encompass the legitimacy of democratic process: laws in the form of European regulation.

From an economic viewpoint, the political decisions makers were optimistic that the international balance between the great powers with regards to the development and trade of technologies would not be compromised to the detriment of Europe. Quite the contrary, they saw the implementation of the embedded mismatch as a potential advantage that would spark renewal and innovation and thus increase global competitiveness. The expectation was that economic growth would be accompanied by more social benefits and social welfare for European citizens.

»Source Text«

“The embedded mismatch will be a virtuous circle that makes Europe the leader in AI technology and preserves a free and democratic society. And indeed, it is only if Europe is competitive in AI development at a global level that we can broadcast our liberal and democratic approach internationally.”—»The European Minister of Global Affairs, Elitist Movement

“It is only if we escape the feedback loops of AI systems, which not only trap people in their uncritical following of mainstream perspectives but also neglect the margins of society, that the real voice of the people will be heard and a system of justice and fairness can rise.”—»The European Minister of Social Affairs, Populist Movement

THE GENERAL REGULATION ON DE-OPTIMIZATION OF AI SYSTEMS

After an extensive public discussion and consultation process, the like of which had never been seen before, the European Parliament passed the »General Regulation on De-Optimization of AI Systems (GRDO) in 2037. The regulation dictates that every AI-application developed, imported, exported, and used in Europe must incorporate an embedded mismatch. It overruled all prior regulations on AI that might have contradicted the new vision, including the »General Regulation on Trustworthy AI (GRTAI) from 2022.

»Source Text«

General Regulation on De-Optimization of AI systems (GRDO):

Chapter 1, General Provisions, Article 1, Subject Matter and Objectives

(1) The embedded mismatch is not meant to stifle AI innovation in Europe but instead uses mistakes, coincidences, and the unforeseen as inspiration to develop a unique brand of AI, one that seeks to protect and benefit both individuals and the common good.

(2) The embedded mismatch accepts and honors the fact that humans are characterized by deviations from logic and the norm. A liberal and democratic political space should enable and support these human traits in order to let European citizens retain their competences, gain new experiences, and develop further skills.

(3) The embedded mismatch aims at protecting Europe from AI systems that determine our future and thus our potential to flourish. Our liberal idea of a non-predetermined and free future should be preserved. This will allow Europe to position itself as a leader in cutting-edge, secure, and ethical AI. Only by ensuring that AI systems differ from a logical and optimized prediction will European citizens fully reap AI's benefits.

In 2038, the powerful »Embedded Mismatch Control Authority (EMCA) went into operation. It aimed to make sure that all the standards associated with the incorporation of embedded mismatches would be fulfilled. This has

been the prime goal of European critical technology assessment since then. Naturally, an embedded mismatch should not completely paralyze systems or harm people; it should challenge the individual user and the entire society in a positive way. For this reason, the GRDO incorporated »The Three Laws of Mismatch.

»Source Text«

The Three Laws of Mismatch:

- (1) The imbedded mismatch may not injure a human being or, through inaction, allow a human being to come to harm.
- (2) An embedded mismatch may not severely damage the functionality of the overall system in which it is incorporated, except where this would conflict with the First Law.
- (3) An embedded mismatch is mandatory and it shall not be revealed as a mismatch, except where this would conflict with the First or Second Laws.

The entire GRDO is based on a set of »ethical principles enshrined in the »European treaties and the »European Charter of Fundamental Rights; it is a canon of values in a European tradition of thought. The principles encompass crucial values such as dignity, responsibility, solidarity, democracy, and sustainability. These ethical principles and the embedded mismatch entered a mutually beneficial relationship: since then the ethical principles have guided the development of AI while the function of the embedded mismatch has helped the European value system to truly unfold. The critique that the Enlightenment and rational thought would be in danger—as brought forward by some Kantian ethicists—was soon defused. Now, it has become the accepted narrative that it is the embedded mismatch that allows our ethical values to unleash full effect.

THE SUCCESS STORY OF THE GRDO SO FAR

Since the GRDO was passed two years ago, AI development has boomed in Europe. A market worth several billion euros around embedded mismatch applications has emerged. There is a broad variety of different functions

and sophistication levels, including major competition between market players in turning predictions and probabilities upside down, developing ingenious »bandit algorithms, and finding the right balance between »false positives and »false negatives.

»Source Text«

“Now, when listening to music, the digital assistants may not suggest music based on past listening habits or other people’s experiences but propose something truly different from the person’s taste so far. The same goes for all other kinds of entertainment and lifestyle products, like films, books, fashion, furniture, accessories, food, or drinks. Mapping services provide people with different routes where they could see or experience something unseen and inspiring, even in familiar settings. If someone is looking for a date or a serious relationship, they may now be suggested a person who would have never popped up using conventional optimized matching algorithms. In classical recruiting processes, supporting tools regularly recommend potential candidates who would have fallen through the net otherwise—like people from groups that had been the subject of discrimination in the past. On the other hand, people searching for new opportunities have been recommended possible projects they would not have thought of before, which opens up unimagined possibilities.” – »Director, Embedded Mismatch Control Authority (EMCA)

As in the past with security checks at the airport, where the alarm regularly went off when nothing suspicious was happening in order to keep the security staff attentive and motivated, people in responsible roles are now confronted with false alarms, false nonalarms, and even absurd suggestions from their support systems in order to prevent a loss of competence and awareness. For example, legal advice services would first recommend that a judge sentence a defendant in a certain way, only to draw attention to the fact that the recommendation was random if the judge was about to follow the recommendation. Medical diagnostic systems would make misdiagnoses on a regular basis so that medical doctors would not blindly trust the system and completely lose important medical skills. If the physician did not notice

the error herself, she would be notified of it by the system and would receive a knowledge refresher package. When it came to therapeutic measures, the system also provided suggestions beyond classical orthodox methods, i.e., traditional Chinese or homeopathic methods, in order to confront medical doctors with alternative points of view. The predominance of humans in other high-profile professions remains of utmost importance because of their guiding and caring function for society.

Beyond this, the *»traditional job market* has changed tremendously. The so-called “3d” jobs—dirty, dull, or dangerous work—have finally vanished. The focus is now on social and creative activities, so called “3f” professions that are fulfilling, fascinating, and fair. The added value creation by the European AI market is so enormous that discussions on salaries and income have become superfluous. The GRDO has made the modern *»post-capitalist age* possible, an age in which individuals are no longer driven by a desire to create wealth but by a desire to contribute positively to the common good by engaging in self-actualization and self-reflection. By making technology human-centered by design, European society has become more open, diverse and tolerant.

»Source Text«

“Nobody—not even the avant-garde³⁴—could have foreseen the tremendous success of the embedded mismatch. We have professionally organized contests on who could live the longest with a voluntarily upgraded high-level embedded mismatch that ‘sabotages’ smart homes, navigation systems, and appointment calendars on a regular basis. Daily life has become a bit of an adventure playground for everyone, but some people even deliberately make it a type of continuous survival training. Our society is bubbling with new ideas that make everyone thrive.”—»Spokesperson, European Research Group on Good Sociological Practice

Crucially, at this point in history, through de-optimization via the embedded mismatch, European citizens have retained the knowledge necessary to be able to deal with life should technology fail. The fears of cyber-attacks did not come to pass. Quite the contrary, while other international power blocks have turned out to be vulnerable to unforeseen technical breakdowns due to a lack of resilience (e.g., *»The US over-optimization incident*, *»China’s*

over-scoring scandal, »Russia's drone tragedy), European people have preserved their main competencies to find their way in analog life. With the help of the embedded mismatch, Europe has now established a new common narrative of true unity, political stability, and freedom.

ISABELLA HERMANN

BERLIN-BRANDENBURG ACADEMY OF SCIENCES AND HUMANITIES

Isabella Hermann is a political scientist and discourse analyst. She explores how artificial intelligence is framed in the discourse and how it affects society. Having always been fascinated by the social effects of new technology, she also works on science fiction and

politics. Currently, Isabella Hermann acts as research coordinator of the interdisciplinary research group “Responsibility: Machine Learning and Artificial Intelligence” at the Berlin-Brandenburg Academy of Sciences and Humanities in Berlin.