

# Profiling with confidence

Big data is booming. Artificial Intelligence allows us to analyse the past, interpret the present and predict the future. More knowledge, more control, fewer risks, and no surprises – a mindboggling cocktail, a dream for politician and businessman alike. The Facebook-Cambridge Analytical scandal rudely disturbed this dream. Suddenly the public became aware that big data technology may be used for good or for evil, depending on who uses it and to what purpose.

So, what is our purpose? Do we want to eradicate crime and disease? Live comfortably? Maximise freedom? Yes, to all! But problems arise when one ethical value must be given up to keep another. Is an ounce of human rights worth a pound of eradicated crime? Can we know what is right? If we do, do others agree?

Ethical guidelines for AI exist. The problem is in knowing how to apply them. We have organised society along the lines of individual responsibility. But AI is not created by individuals. It requires complex collaboration of professionals across barriers of time, space, expertise, and command chains, serving visible and invisible stakeholders. Our challenge is to find a way to do ethics together. So that teams, not individuals, may produce trustworthy AI.

Imagine this newspaper article:

## SHELTERED HOME DRAMA ROBBED OF SAVINGS BY FAKE INSPECTORS

Big data analysis showed that the swindling might be related to the Gimbodian community. The Dutch Tax office then tagged all Gimbodians for income tax checking. Some Gimbodians turned out to have no income, apparently living on nothing, which was considered suspicious. The departments of Finance, Social Security and Justice subsequently coordinated the arrest of a gang of Gimbodians. Before the case got to court, public opinion turned against the Dutch Tax Office on account of discriminating against Gimbodians; the other departments were slapped for acting on improper evidence. The State Secretary for Finance took the wrap and stepped down. In a joint statement, the ministers of Finance, Social Security, and Justice apologised profusely. The IT department was blamed, because ethnic profiling should have been impossible ‘by design’.

Gimbodians do not exist. However, substitute any adjective, such as red-haired, holding a double passport, unemployed, or female, and the story will be the same. If you belong to a minority heavily prejudiced against in the past, ‘big data’ analysis will show you to have less than average money, education, stability, space, and happiness. As a statistic, you will have more psychological problems, debts, criminal convictions, and health problems.

Yet by itself data does not judge or discriminate. Data profiles tell the story of how things have worked out in the past. The question is how to use this knowledge. We might say that the end justifies the means: vulnerable people must be protected, so profiling potential criminals is fine. Or we might argue that it is not fair to suspect someone of swindling on account of being Gimbodian, and therefore profiling is out.

This is about ethics. There may not even be a correct answer. And even if we find the right answer, will everyone agree?

The EU AI Expert group has also been struggling with this question. They suggest the use of four ethical imperatives which are firmly grounded in European Law: *respect for human autonomy, prevention of harm, fairness, and explicability*, in that order. However, these principles may conflict, as in our Gimbodian example. Pandemics and acts of terrorism have also shown us that that *prevention of harm* and *human autonomy* may pull us in opposite directions. To offer even more practical guidance to the AI practitioner, the EU AI Expert group has published the [Ethics guidelines for trustworthy AI](#)<sup>1</sup>. The challenge is in applying these guidelines. “The problem of homo sapiens is not to formulate ethical rules, it’s to put them into practice<sup>2</sup>”.



## Research brief CIO Edition

### RECOMMENDATIONS

- ▶ Chief Information Officers that have dealings with big data – and who has not? – should establish a corporate code for its creation and use. This code must contain three chapters: aspirational, advisory, and disciplinary. For European companies, the aspirational chapter should be aligned with European AI proposal; the advisory chapter should provide instructions on how to implement ethical guidelines for trustworthy AI into professional working routines; the disciplinary chapter must state specific rules to be followed always.
- ▶ Project management methods need extending with ethical considerations and with new stakeholders such as society, vulnerable groups, and future generations. CIOs may create this extension themselves. Alternatively consider collaboration with market consortia, such as The Open Group<sup>3</sup>.
- ▶ When integrating ethical risks in the company risk framework, consider appointing a Chief Ethics Officer. This is a new CxO function on the rise<sup>4</sup> specifically to deal with ethics-based risks and balance these against the need for a company to maintain a competitive edge.

### TRUST THE TEAM

There is a danger of taking the [Ethics guidelines for trustworthy AI](#) as a to-do list, and to use it for playing the ‘blame’ game. To illustrate, let us go back to our newspaper article, where IT department did not prevent the selection of Gimbodian suspects. Indeed, they had not prevented *unfair bias*, nor had they paid attention to *societal wellbeing*. Does that make them blameworthy? Probably not. For someone to be blameworthy, we assume<sup>5</sup> that this person:

- did something wrong and this ‘wrong’ action contributed to the ill effect;
- was free to choose the ‘wrong’ action;
- could foresee the consequences of the ‘wrong’ action.

IT departments and engineers tend to do as they are told. In this example, they were presumably neither free to take another action (b), nor in a position to foresee the consequences (c) of their actions. A similar argument holds true for the politicians involved. They may have been free to choose their action (b), and even to foresee the consequences (c); however, they did not commit the actual action that contributed to the ill effect (a). Clearly there is collective responsibility for the end result, but it is difficult to translate this to individual responsibilities.

This phenomenon is known as the “many hands” problem<sup>6</sup>. It is typical for high-impact technological advances, such as the creation and responsible use of big data. The situation requires expertise from professionals from widely divergent backgrounds: technical, legal, psychological, and political, to name a few. They usually belong to different organisational and administrative structures. They contribute at different times and to different stages of the project.

### CONCLUSIONS

‘Big data’ comes with many benefits but also with big responsibilities. When we use historic data to interpret the present and predict the future, we may introduce bias, injustice, and tunnel vision. How to avoid this? Not by leaving the issue to government or to law<sup>8</sup>. Or by playing the blame game. Certainly not by waiting for more technology. The EU AI Expert Group has created ethical guidelines for trustworthy AI – basically, a list of ethical issues to think about in every AI project. The challenge lies in incorporating these guidelines into the fabric of organisations and teams. This will not happen by itself, but will require much effort and support from the corporate level. The tone at the top is of vital importance. An excellent opportunity for the CIO platforms Netherlands and the CIO Dutch government to lead by example.

In effect, the members of the big data team are separated by time, place, knowledge and skills, and yet have to work together in creating a collective result that stands up to ethical scrutiny. Organising this kind of teamwork is not trivial. Project management methods strive to cut out doubt and uncertainty in favour of completing on time and within budget. That is why goals are set in SMART<sup>7</sup> terms: Specific, Measurable, Achievable, Realistic and Timely. In contrast, the *Ethical guidelines for Trustworthy AI* are any project manager’s nightmare. They introduce tough questions; cause uncertainties to keep popping up and generally have the potential to redirect the entire project.

Here we see the implementation gap between the *Ethics guidelines for trustworthy AI* on the one hand, and the toolkit of the typical project manager on the other. Only top-down corporate action can fill this gap.

- One way is by taking a leaf from risk management. If an ethical AI guideline is not followed, something or someone will be at risk. It should be crystal clear what risks cannot be decided within the project, but should instead be taken to the corporate level. This would apply to *Societal impact*, for instance.
- Another way is to develop a company standard implementing some of the more practical ethical guidelines. This would apply to *transparency* and *accountability*.

Whatever way is chosen, consider taking one step at the time. This is not just about instrumental change. It is also about finding new ways to ensure collective responsibility. About changing the hearts and minds of organisations.

### References

- 1 European Commission. (2019, April 8). Ethics guidelines for trustworthy AI, see <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>
- 2 Haselager, P., & Mecacci, G. (2020). Superethics Instead of Superintelligence: Know Thyself, and Apply Science Accordingly. *AJOB Neuroscience*, 112, 113–119. <https://doi.org/10.1093/ajob/nqaa033>
- 3 [www.theopengroup.org](http://www.theopengroup.org)

4 Forbes. (2020, June 29). *Forbes Insights: Rise Of The Chief Ethics Officer*. See: <https://tinyurl.com/y362mwy>

5 Poel, van der, I., & Royakkers, L. (2011). *Ethics, Technology, and Engineering*. Wiley-Blackwell.

6 Wikipedia, see <https://tinyurl.com/yag9g2>

8 Sloot, B. van der, & Schendel, S. van. (2019). *De modernisering van het Nederlands procesrecht in het licht van big data*. Retrieved from <https://tinyurl.com/y95vsvp>

### Colofon

This research brief has been compiled by Inge Wertwijn. She is an independent privately funded researcher as well as a Dutch government employee. In constructing this research brief, the author has not used confidential information from governmental sources. The position of the Dutch government on this issue is not known, and therefore has not influenced this text.