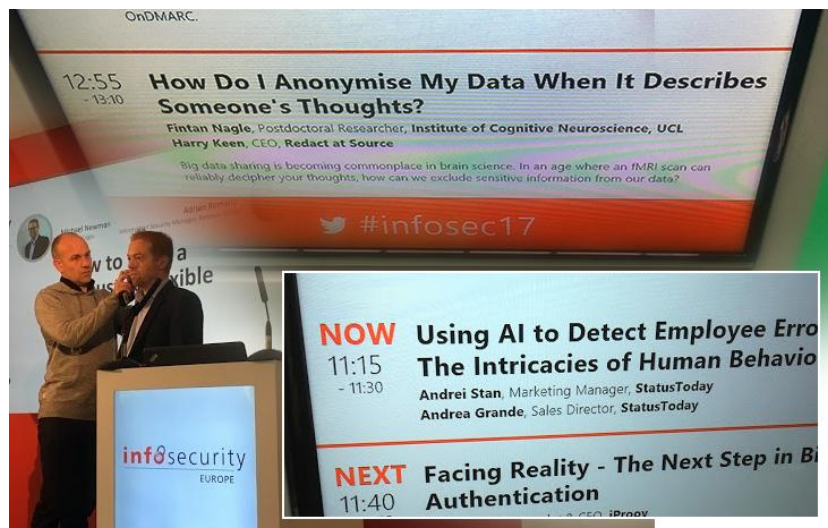


Can AI save struggling companies from Europe's new privacy laws?

Originally published at <https://www.intelligentautomation.network/artificial-intelligence/articles/why-ai-probably-cant-save-laggard-companies-gdpr>



Companies are turning to AI to protect themselves from crippling European privacy fines - but can Machine Learning solve a problem which no-one has really defined yet?

Security and compliance vendors are partying like it's 1999; that was the last time any pending digital threat offered such lucrative possibilities as the European Parliament's General Data Protection Regulation (GDPR), which comes into force in May of 2018.

But Y2K II is a surer bet than the no-show millennium bug. Non-compliance risks a €20m fine or 4% of a company's annual turnover, whichever is greater.

Many UK companies waiting out the terms of Brexit seem to have procrastinated their way into a potentially costly situation, which vendors have responded to with a wide range of AI-driven solutions — offerings which also have to be careful of coming into contact with Personally Identifiable Information (PII).

Analysing 'anonymous' targets

"For one of our large clients," says Luke Goldspink of [StatusToday](#), "we can anonymise the data before it comes to us so that we actually have no PII. We might just get a string of letters as a name or an email address, and we don't know who it is. The company will, but we won't."

Though it's debatable whether an email address is suitably 'anonymised' information, promotion for the systems on offer do emphasise their ability to analyse data without identifying individuals.

The supervised learning systems on display all seem to be dealing with hashes, IP/email addresses or other kinds of tokens, rather than 'people' — unnamed entities logging back into systems they were long-since fired from for a spot of recreational exfiltration; or setting three-month launch dates on a malicious Cron job as the security guard approaches with a cardboard box and a steely gaze.

But in all cases the malice is anonymous, the 'identity' in digital escrow until the report of the culprits' actions are returned to the client. Though the vendors emphasise GDPR compliance as a marketing tool, it's also in their own interests, since the new rules will apply equally to them.

Trouble in the mountains

GDPR brings new urgency to the issue of the legacy data mountains which are maintained by so many larger or older companies. All that unstructured data will be just as susceptible to the new regulations as more current or active data streams. Perhaps the stick of regulation will force the action that the carrot of Big Data could not.

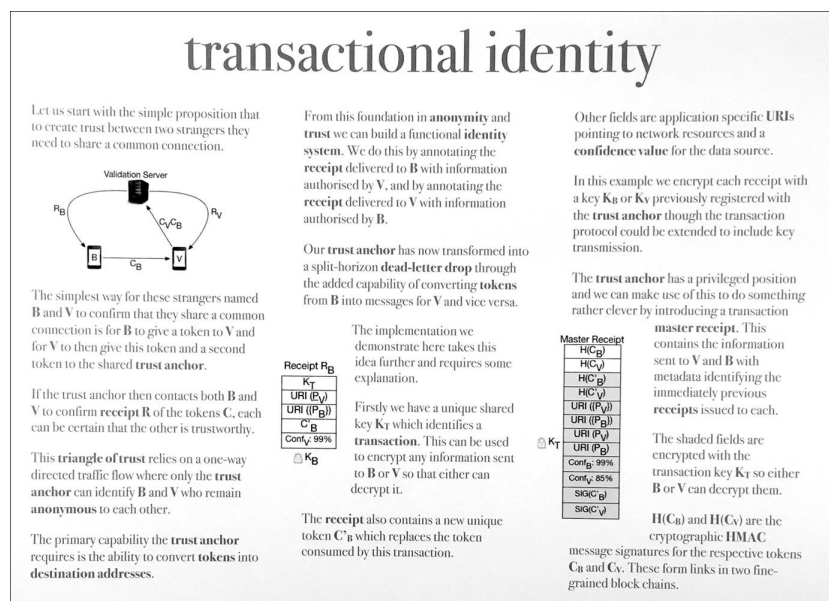
This is the stock-in-trade of Berkshire-based startup [Exonar](#), which specialises in data discovery, appending metadata and classifications to otherwise abstract documents based on their content. Chief Operations Officer Julie Evans has found a significant spike in enquiries related to GDPR recently.

"It's a complete hockey stick" she says. "A huge number of CIOs and CISOs will say 'My biggest problem is that I don't know what's contained in the information storage I've got.' I recognise that problem...that's the thing that people have been saying most often, the thing that 'keeps me awake at night'."

Former physicist Eleanor McHugh, of London-based software consultancy Innovative Identity Solutions, agrees with me when I suggest that the lack of scalability of so many solutions proposed at InfoSec this year is part of the 'identity problem' which GDPR has made critical: most of these schemes could only ever work with 100 per cent national adoption, from the stalls to the winning line in one unfeasible bound. McHugh also believes that the one the UK government [is building](#) is also too fractured to be truly promising.

"We know some of the people at Verify," she says. "and to be honest, they're going nowhere. They're actually moving backwards in time, if anything. The design for Verify could be simplified to just be the system that we've got up there..."

She points to the mounted explanation of a trust system recently commissioned by a client:



(Simplicity is relative, of course.)

"It would work really well. Everybody who's trying to implement Verify seems to be moving backwards with tokens and FIDO and back to [Public Key Infrastructure]. It's easy, familiar — and it doesn't work."

McHugh suggests that the commercial push to profit from transactions is part of the reason that most of the proposed schemes don't prove to be viable.

"You can't charge a transaction tax - that won't work." she says, suggesting that individuals be charged to write a new profile, rather than trying to make the framework free and movement within it chargeable. "They all see the dollar signs around the transaction - the small dollars in the profiles and the big dollars in the transactions."

Exact solutions for an ill-defined problem

It's no surprise that vendors are leveraging AI in their solutions to the GDPR challenge; notwithstanding the current cachet about it, Machine Learning has real potential for the kind of pattern-detection needed to enable compliance.

However, it needs a well-defined template to look for, and this is a problem. At the recent cybersecurity expo InfoSec Europe, one of the many frantic topic headlines (see illustration, top) was 'How Do I Anonymise My Data When It Describes Someone's Thoughts?'. There are many similar problems involved in actually defining PII before the search can even begin. It's possible that the European Council has made this rather philosophical issue urgent many years before AI is ready to address it. The medium is binary, but identity, as a property, remains a kind of abstract fog.

It's ironic that just as the populace is becoming increasingly compliant and willing to adhere to the nearest and best-diffused ID system (usually Facebook's openly-available login API, or else one of the East's [more Draconian 'Real Name' schemes](#)), the rapacity of corporate data miners has forced a crisis around such a nebulous enemy.

Illustration: Martin Anderson