The Validation Crisis

Brenno de Winter

Brenno@dewinter.com, +31653536508

Brenno de Winter

- Born in 1971, wrote first program at age of 5
- My themes are information security and privacy
- Former investigative journalist/Journalist of the Year 2011
- Winner of the 2012 Internet Society Award
- Hacked Public Transportation Card
- CSPO during Corona-pandemic for the Kingdom of the Netherlands (Netherlands, Curacao, Aruba, St. Martin)
- Spiritual father of OpenKAT and the Methodology for Information Security Research with AUditValue (MIAUW)
- Wrote multiple books
- Investigated 2,000+ security incidents
- Like cats, boats and cooking

Ohh no ....
There will be
AI involved in
this talk

What is the validation crisis

- The validation crisis is the phenomenon in which organizations and individuals act on incomplete, unverified or erroneous information
- Observation: Decisions are often made without sufficient validation of data with sometimes dire consequences
- AI and automated decision making automate this problem to greater magnitude
- How do we arrive at and address the validation crisis?

Example: a stalker

- An intern at a mental health facility files a report against a patient.
- There is stalking with text messages
- The woman can show the police the messages
- The messages are threatening and racist in language
- The man is arrested and denies
- The judge finds the evidence legal and convincing
- The man is convicted and jailed for one year

A stalker …
after his release

- Does the stalking begin again
- Victim turns to Thijs Zeeman
- Investigation reveals things are not right
- Victim turns out to be based on spoofing
- What makes the chain so failed?
- Demanding traffic data

A woman shows up at the general practicioner

- Medical field.
  - Misdiagnoses due to routine: When doctors rely on routine without thorough examination, it can lead to fatal errors.
  - Example: Women with heart disease whose symptoms are often taken less seriously, which can lead to delayed treatment.

Some causes of the Validation Crisis

- Technology dependence.
  - People blindly trust technologies such as AI and automated systems without checking them thoroughly.
  - Example: The benefits affair where algorithms falsely labeled parents as fraudsters.
- Cognitive biases.
  - Confirmation bias: People seek confirmation of what they already believe rather than the truth.
  - Example: Management teams rely on gossip or vague signals without thorough verification.
- Inadequate checks in automation processes.
  - AI and automation systems can make mistakes if they are not properly trained or checked. Lack of control can lead to serious problems.

Comes a banker in Singapore

Financial sector: Nick Leeson and Barings Bank: Lack of supervision allowed a single trader to bring down a bank.

Techno-optimism

- Belief in Unlimited Potential: Techno -optimism refers to the belief that technology can solve most, if not all, of humanity's problems
- No regard for risks
- Overturning critical voices
- Reduced sense of reality

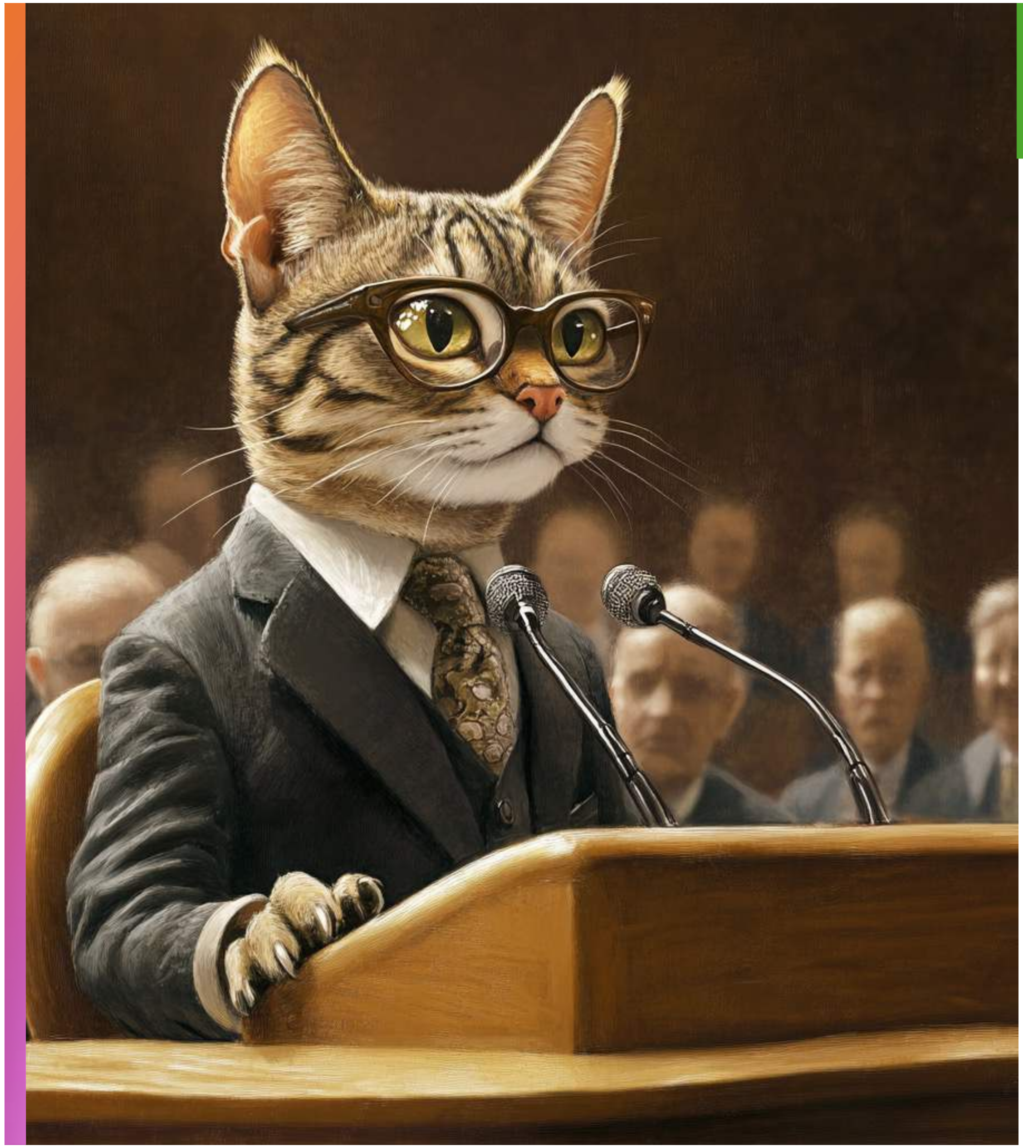
Computer says no

- Definition: The phrase “Computer says no” refers to situations in which computer systems make decisions without human intervention, and those decisions are not challenged, even if they are incorrect or unreasonable.
- Example: Government agencies automatically withdrawing benefits based on erroneous data without human intervention, as in the benefits affair.
- Consequences: Can lead to major human and legal errors if there is no room for human review and validation.

Schnapsidee

Allowances affair

- The Bulgarian fraud
- Dutch government claimed back thousands of euros in benefits from parents based on automated decisions
- There was no adequate validation of data, and human circumstances were not taken into account
- Thousands of parents were wrongly accused of fraud, leading to financial and social disruption
- Human dimensions and validation are often lacking in complex, automated processes, leading to serious errors



Schnapsidee

Boeing 737 Max: flawed validation

- Remote Management
- Relying Too Much on Technology: Boeing introduced a new automation system (MCAS) without sufficient pilot training or thorough testing.
- Reliance on Assumptions: The company assumed pilots would react quickly to system failures, but these assumptions were not validated in realistic scenarios.
- Errors in validation process: Internal reports about system risks were ignored. Crucial safety checks were missing in the rush to get the aircraft to market.
- Consequences: Two fatal crashes (Lion Air and Ethiopian Airlines) and global grounding of the 737 Max, with thousands of lives affected.
- Lesson: Validation processes must be complete and unbiased. Assumptions without thorough testing can have catastrophic consequences.

Not validating
a bad and expensive joke

- Online scams in U.S. estimated at $8.8 billion
- The affair with Boeing
  - Damage 737-Max affair
    between $20 billion and $30 billion
  - Recovery of the company: $25 billion
- The surcharges affair:
  - Bulgarian fraud: up to 4 million euros
  - Surcharge affair: 14 billion in 2025 and counting
  - 3.500 times as much!

Training AI: a matter of validation

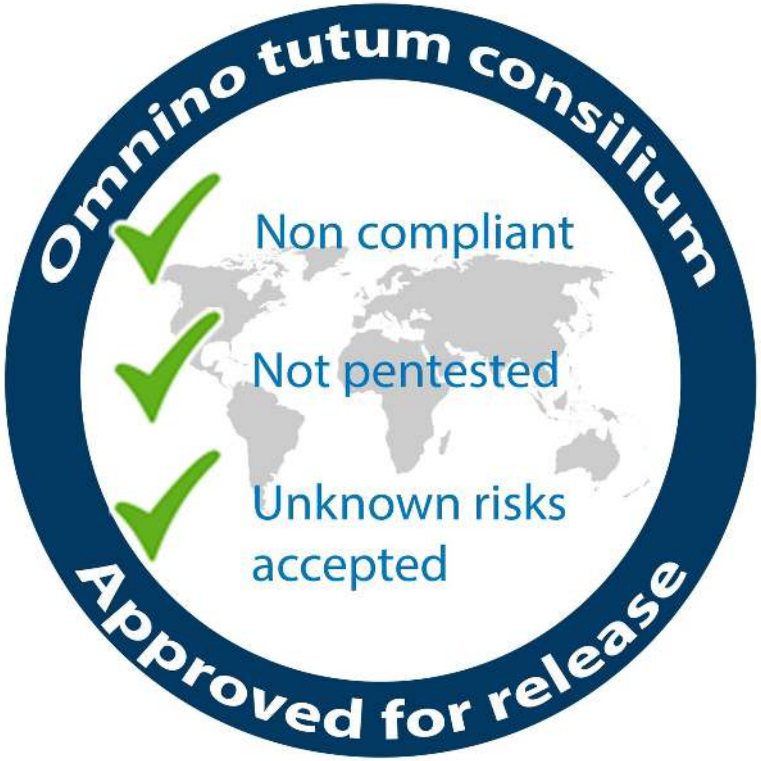
- Data Collection: An AI model learns from data. The more diverse and representative the data set, the better the model can generalize.
- Training and Validation Sets: Data is broken down into training and validation sets. The training set is used to learn the model, while the validation set is used to check how well the model performs on unseen data.
- Overfitting: The model learns too specifically on the training data and performs poorly on new data.
- Underfitting: The model does not learn enough patterns and therefore performs poorly on both training and test data as a result.
- Bias in Data: Unbalanced or biased data can lead to a model making discriminatory or inaccurate decisions. Diversity in data sets is crucial to reduce bias.
- Continual Learning: The model can be periodically updated and re-trained with new data to keep it up-to-date and ensure continued accuracy.
- Validation is Crucial: After training, the results must be continuously validated to verify that the model does what it is supposed to do, without faulty assumptions or unforeseen outcomes.

An example with AI

There is one more more thing

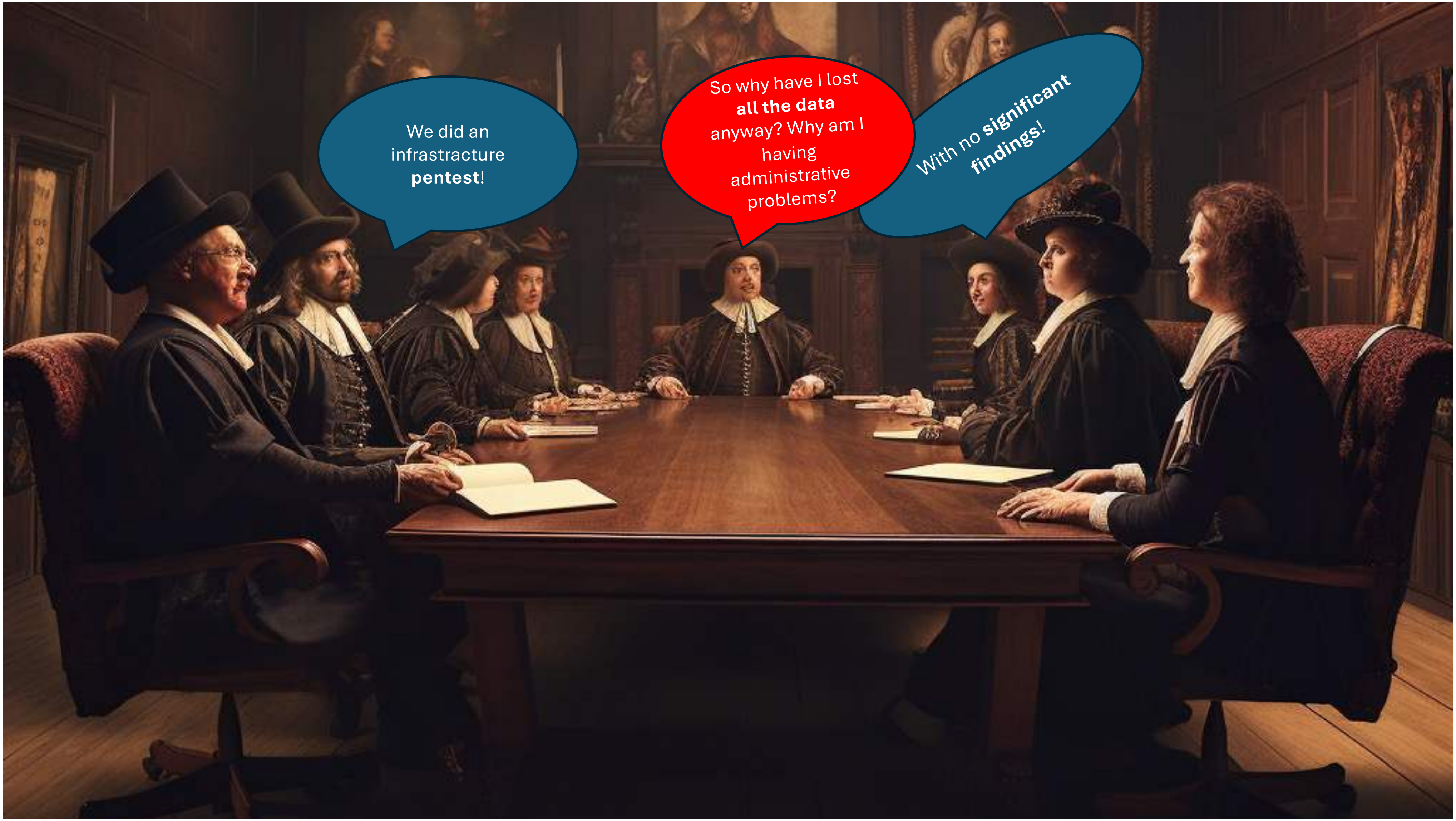
Customer **techno-optimism**: **We are totally secure!**

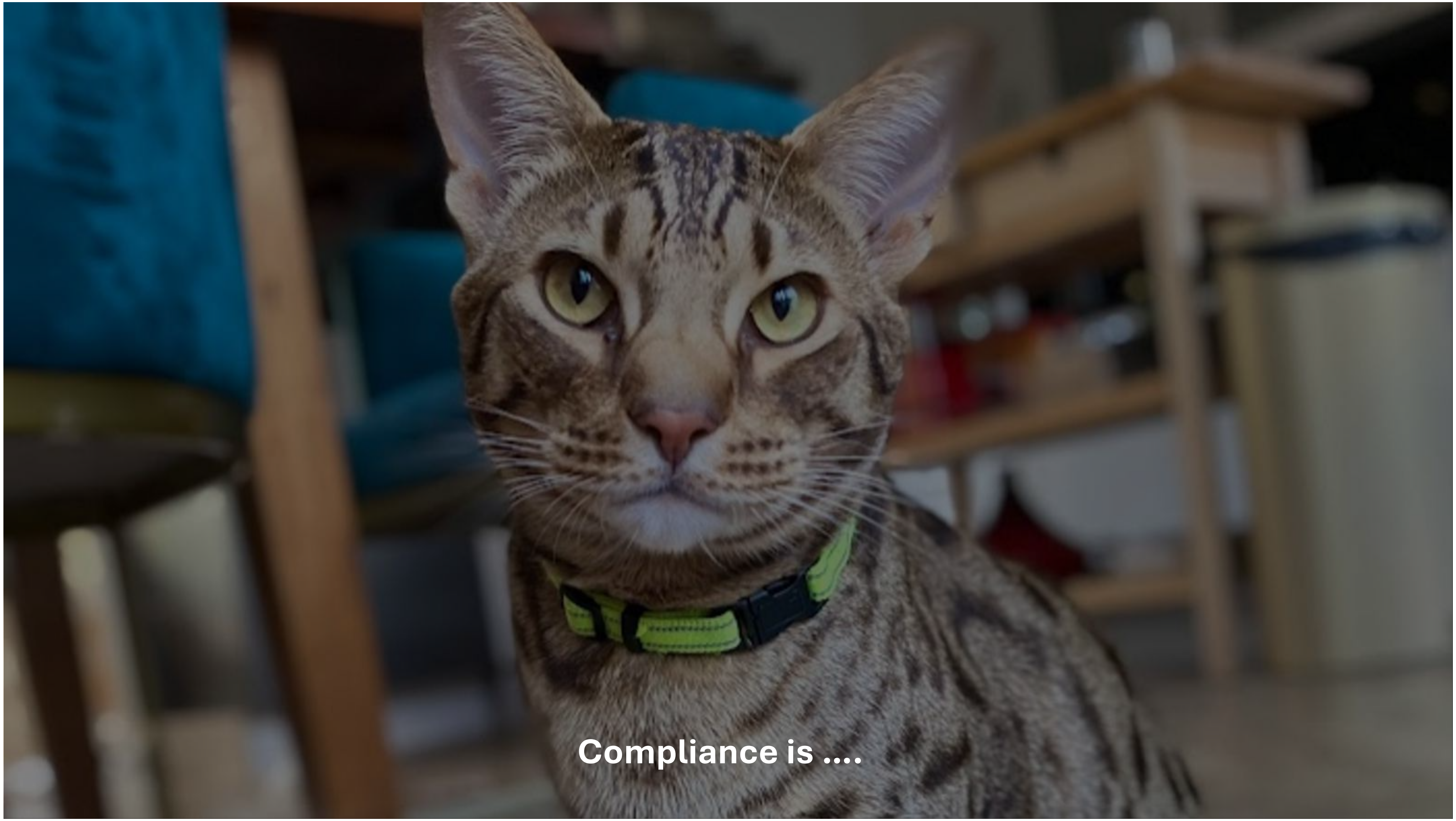
- ‘We've had hackers watching’
- ‘We take inspiration from standards’
- What our people do is magic
- What a pen test is, well that's what we do!
- She couldn't get through
- Even hackers couldn't get it broken



Often this is reality

Six months later ...


We did an infrastructure **pentest**!
So why have I lost **all the data** anyway? Why am I having administrative problems?
With no **significant findings**!



Compliance is ....

The broader question:
How do you conduct an audit?

- Is that testing a claim?
- Is it looking back?
- Is it looking forward?
- Is it being convinced?

Wat do you steer towards?

Repairable with some surgery

- Less magic
- Provide more certainty
- Providing evidence

Examination according to minimum set of tests

- Uniform requirements for procurement of pen tests
- All findings weighed against the same yardstick (CVSS)
- Clear agreements on how to write down findings
- A minimum set of requirements:
  - OWASP MASTG/WSTG
  - CIS Benchmarks
- Uniform way of presenting findings: PTES
- Reproducibility of the study including underlying evidence
- Underlying tests are included
- Pentester puts signature
- Optional: the report must be able to become public

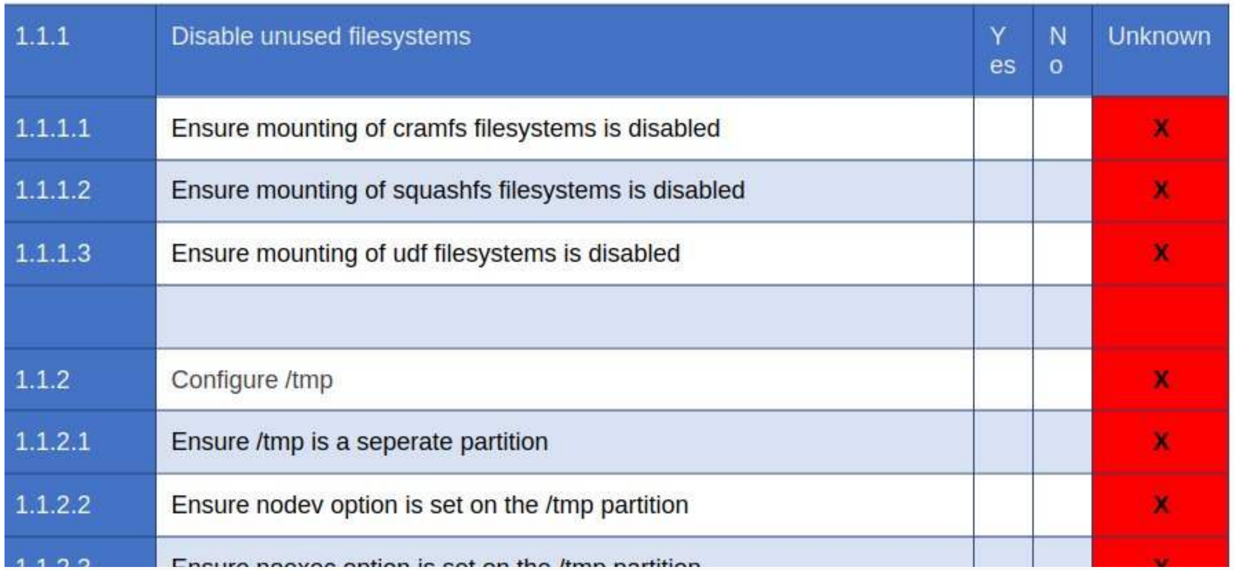
After penetration testing …

- Report of findings by ‘a qualified auditor’
- Affordable provides a summary based on the underlying document
- Provides assurance on the underlying documents

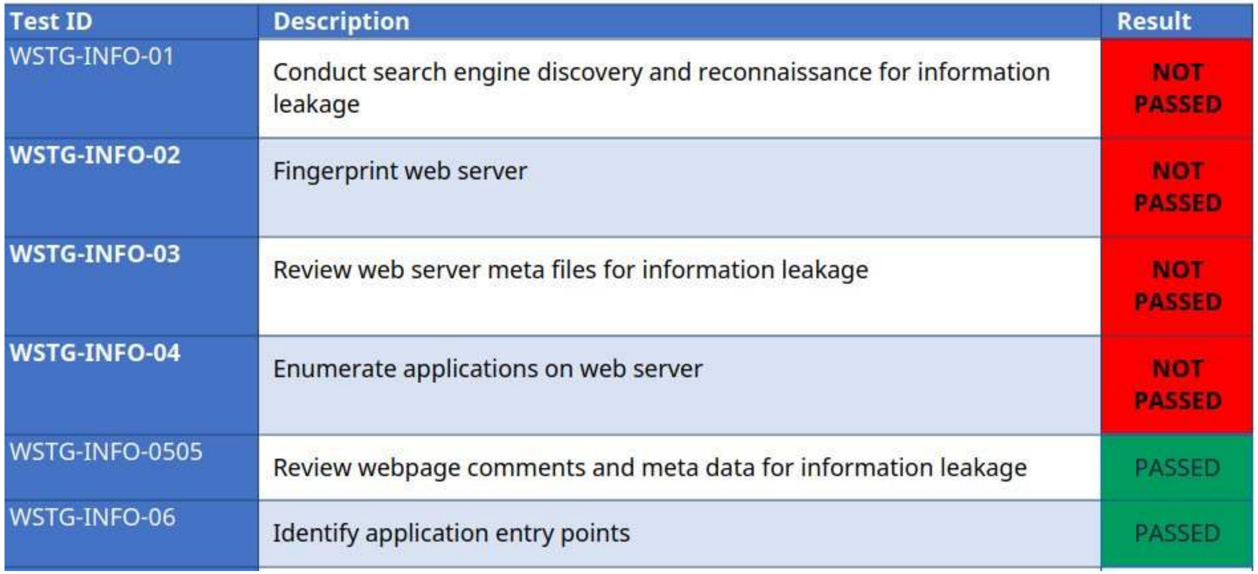
Methodology for **information security research** with **Audit Value** (MIAUW)

An open source pentest standard for all

What is it?



1.1.1	Disable unused filesystems	Y es	N o	Unknown
1.1.1.1	Ensure mounting of cramfs filesystems is disabled			X
1.1.1.2	Ensure mounting of squashfs filesystems is disabled			X
1.1.1.3	Ensure mounting of udf filesystems is disabled			X
1.1.2	Configure /tmp			X
1.1.2.1	Ensure /tmp is a seperate partition			X
1.1.2.2	Ensure nodev option is set on the /tmp partition			X
1.1.2.3	Ensure noexec option is set on the /tmp partition			X



Test ID	Description	Result
WSTG-INFO-01	Conduct search engine discovery and reconnaissance for information leakage	NOT PASSED
WSTG-INFO-02	Fingerprint web server	NOT PASSED
WSTG-INFO-03	Review web server meta files for information leakage	NOT PASSED
WSTG-INFO-04	Enumerate applications on web server	NOT PASSED
WSTG-INFO-0505	Review webpage comments and meta data for information leakage	PASSED
WSTG-INFO-06	Identify application entry points	PASSED

- Set of controls for a pen test:
  - Requirement
  - Description
  - How do you validate this as an auditor?
  - What do you get from this?
  - What are you missing if you don't have it?
  - How do you ask for this in procurement?
- Official procedural report by auditor:
  - Going through the process correctly
  - Overview of the findings
  - Evidence that for the basics there is or is not in control.
- A model pentest report
- Legal information
- An advisory guide on the ‘how to'

Looking at the same

- Clarity about what has been investigated
- Clarity about what you did not get and what that means
- Clarity about whether the research was actually carried out
- Reproducibility
  - **Verifiable**
  - **Not doing the same test more often**