

# Testology and Technology: how the human factor can leverage and enhance the use of AI in making good tests

Nico Dieteren MSc - Dutchtestologist

© dutchtestologist 2025

# Outline

- 1: What makes a test a good test?
- 2: Can AI help?
- 3: Combine AI-models with human expertise
- 4: The African context
- 5: Conclusions and recommendations

# What is the purpose of educational tests?

‘..... they provide information to support decision making....’

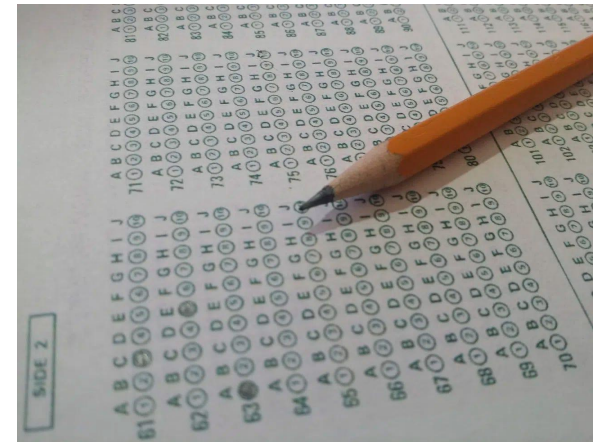
# What is the purpose of educational tests?

‘..... they provide **information** to support **decision making**....’

# Test: a definition

“A test is an objective and standardized method for estimating ability, based on a sample of that ability.”

Estimated ability:  
a gap between observed ability and true  
ability



# Examples of estimating ability by observing behavior

► Strong



► Poor



# Quality of information and decision making

A good test should be:

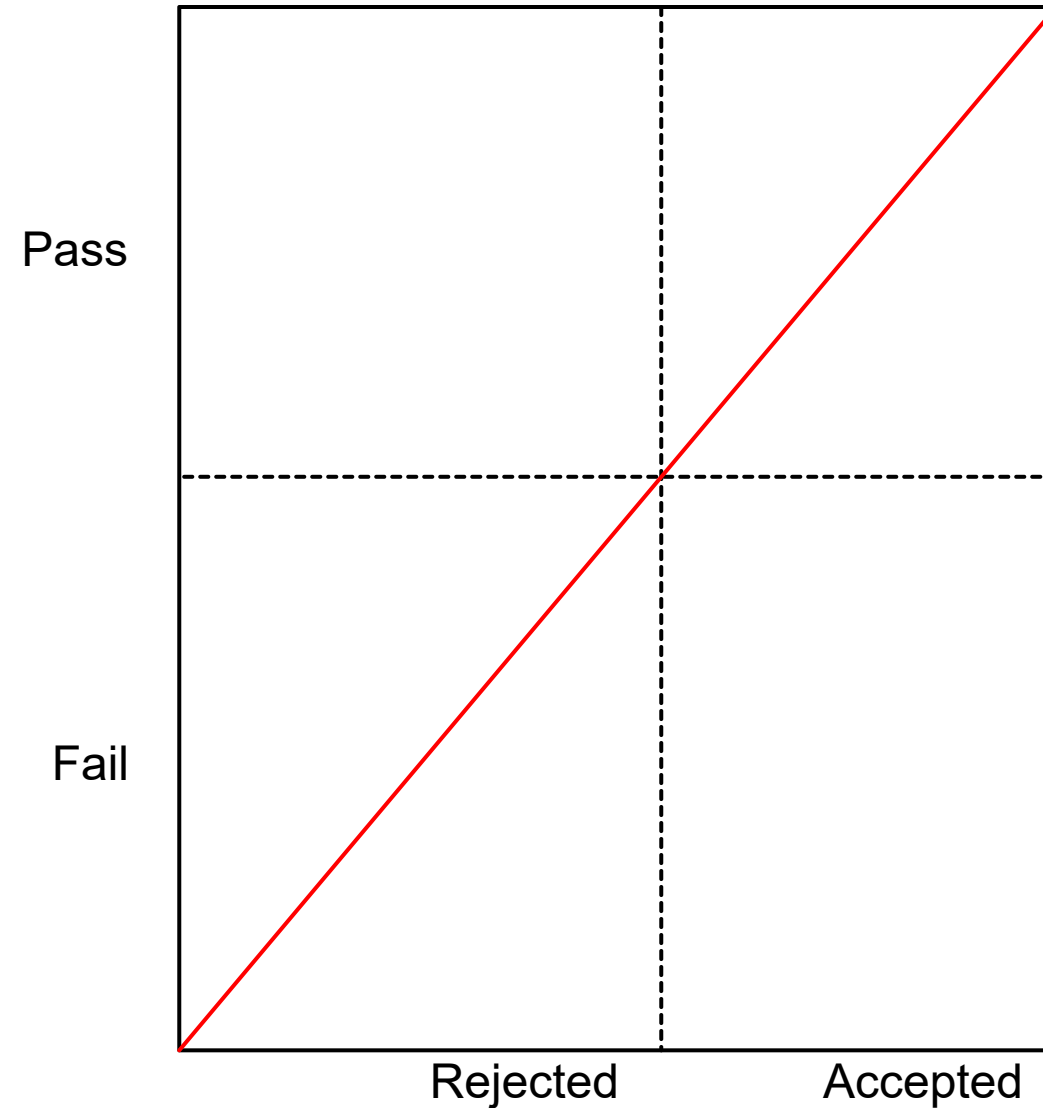
- Valid
- Reliable
- Objective
- Fair



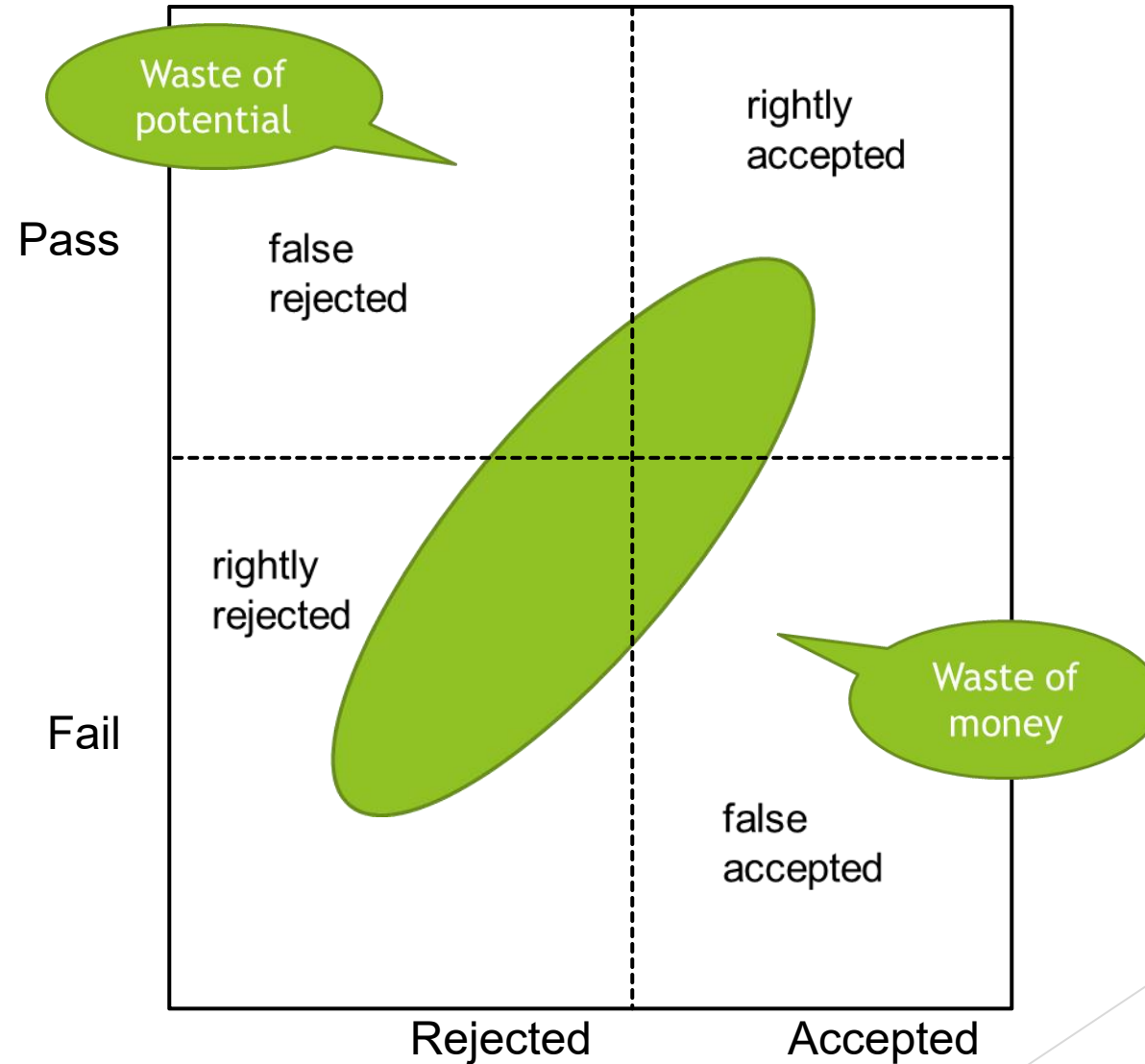
# We Humans are Poor Observers!



Based on observations we make decisions.....



.....some good and some not good.....



Can AI help?

# “Strong at some stages, unreliable at others”

AI can improve elements when it comes to

- standardising procedures
- automatising processes
- higher efficiency for large scale
- providing big quantity in less time
- reducing failures of the human factor

► *For example:*

- *automated item generation*
- *test assembly*
- *test administration*
- *proctoring*
- *marking and scoring*

AI becomes unreliable when it comes to

- creativity for item generation
- ethical judgements about items
- intellectual property rights
- consistency over time
- *high demand for stable energy and water resources*

The background features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, ranging from light lime to dark forest green. These shapes are concentrated on the right side of the image, creating a modern, layered effect. The text is positioned on the left side of the image, set against a plain white background.

Combine AI models with human expertise

# AI models

(www.shriftman.medium.com)

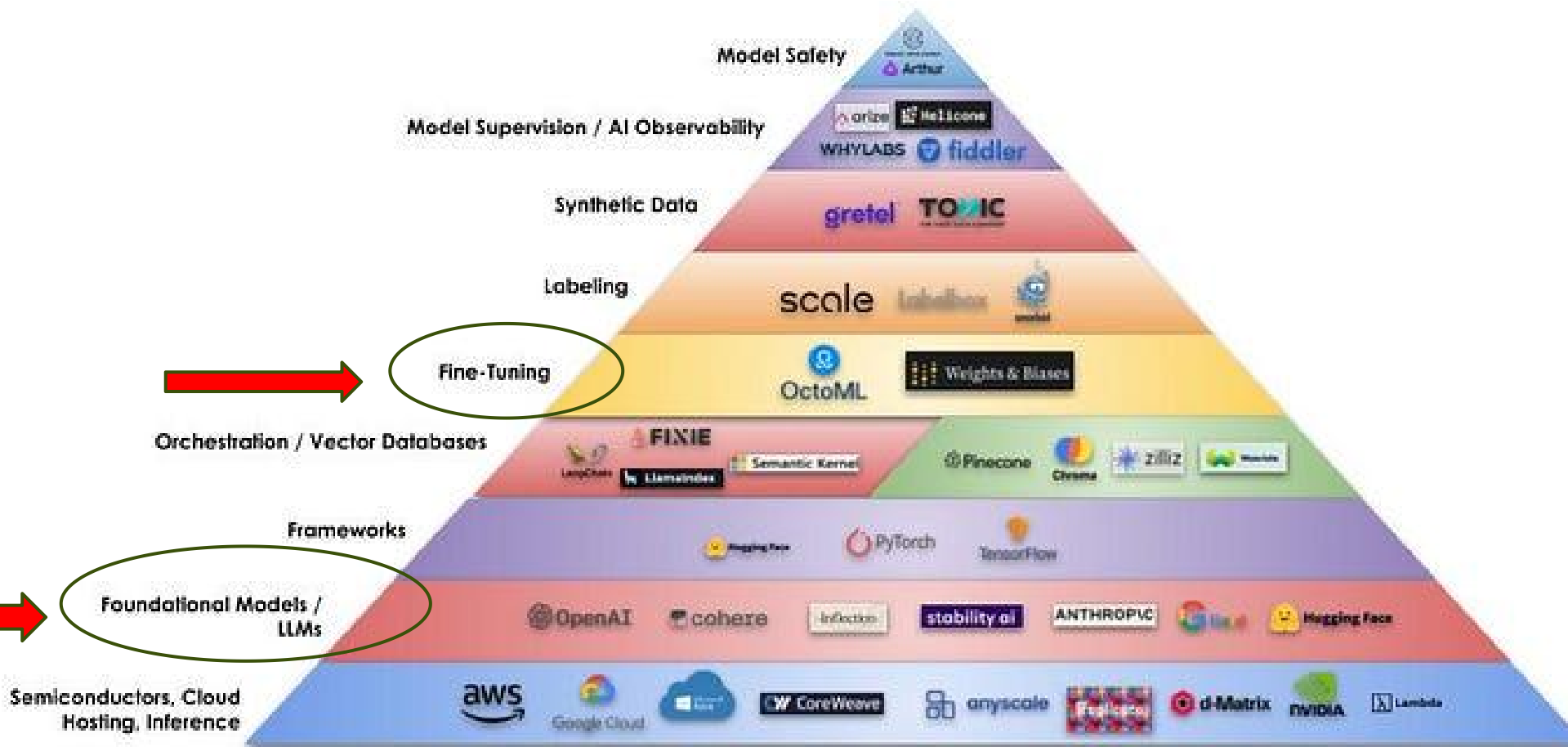
## The Building Blocks of Generative AI



# AI models

(www.shriftman.medium.com)

## The Building Blocks of Generative AI





# General purpose models

- ▶ Foundation for a wide range of applications
- ▶ Large Language Models (LLMs) : to generate new texts on the basis of existing text, using 'deep learning'

# General purpose models

- ▶ Foundation for a wide range of applications
  - ▶ Large Language Models (LLMs) : to generate new texts on the basis of existing text, using 'deep learning'
- 
- + high level of language understanding
  - + pre-trained on vast databases with texts and data
  - + very cost-effective once trained
  - vast data and texts processing in the initial stage: time and money
  - 'hallucinations': incorrect, nonsense or nonethical output
  - 'generalisations': lack of specificity for your purpose
  - high risk of bias for small languages that do not have big LLM database

# Finetune AI models

- ▶ Customisable for a specific purpose
- ▶ Also based on the use of LLMs
- ▶ Model to be trained in fine-tuned domain-specific data
- ▶ Output /responses from the model are aware of context

# Finetune AI models

- ▶ Customisable for a specific purpose
  - ▶ Also based on the use of LLMs
  - ▶ Model to be trained in fine-tuned domain-specific data
  - ▶ Output /responses from the model are aware of context
- 
- + Same high level of language understanding
  - + More specific output, tailored to your purposes
  - + Checks to avoid 'hallucinations'
  - + Less resources needed in the iterative process once finetuned
  - \_ Finetuning needs to be updated regularly: costs more time and money
  - \_ More domain specific human expertise needed

# Finetune AI model: an example

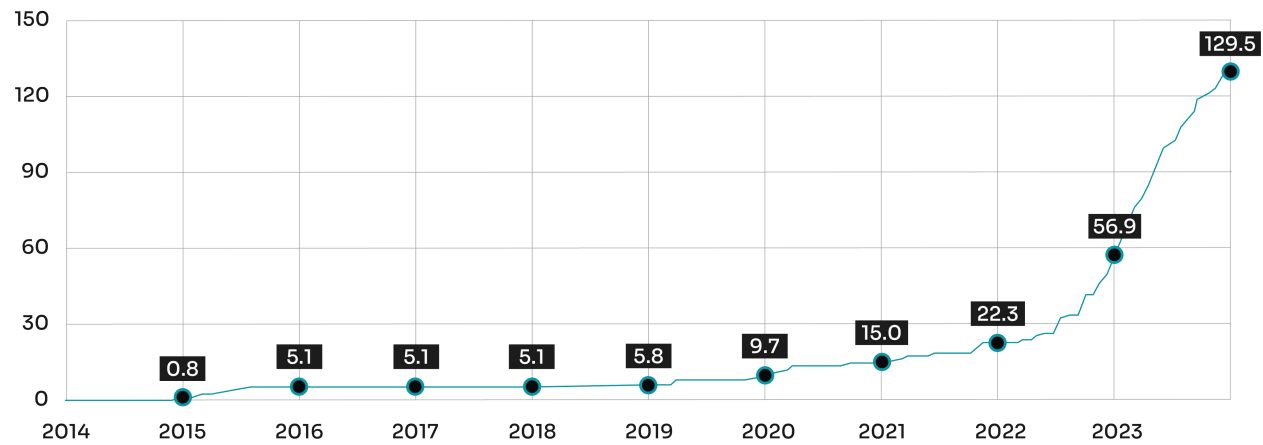


Improve your exam's security with **Finetune's Generate™**, brought to you by Prometric



# The African context

**Blackout time due to load-shedding**  
Accumulated days since January 2014



Data source: [businesstech.co.za](http://businesstech.co.za) / Pieter Jordaan

picture: Esa Alexander/Reuters



# The African context



- ▶ Local/regional initiatives, like Indaba and .....?
- ▶ Avoid to be 'The Missing Continent'
- ▶ But, make sure the use of AI will fit to your culture and ethical standards
- ▶ DO NOT FORGET THE POWER OF THE HUMAN FACTOR

# Conclusion



- ▶ AI can help improve quality of tests and item writing, but the human factor will be needed for checks and fine-tuning.



dutchtestologist@outlook.com



**NICO DIETEREN**  
**"TESTOLOGIST"**

Consultant and Trainer