

# “Are we there yet?” The promise of the Holodeck for the Future of Education



Alina A. von Davier |  
Duolingo  
April, 2023

# The world is changing.



# Education is being **redefined** by



# The Promise of the Holodeck

**Learning & Assessment  
are personalized,  
adaptive, and immersive**



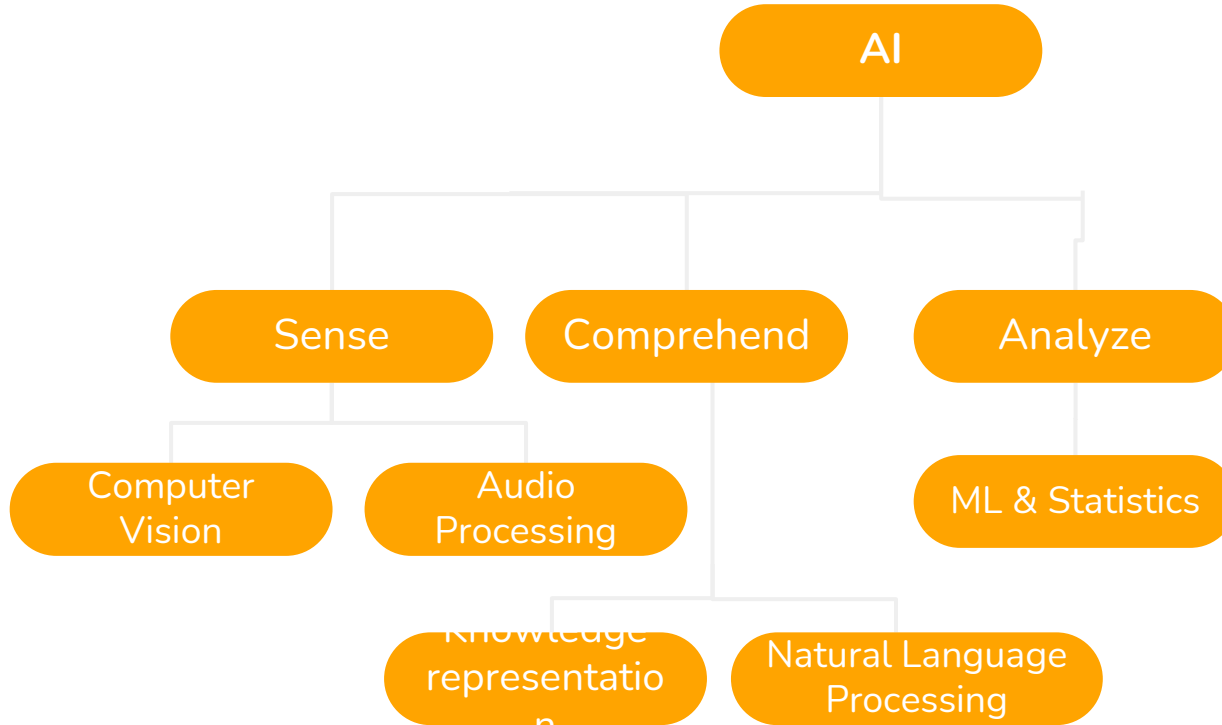
William Riker (Jonathan Frakes) entering a Holodeck simulation.  
Star Trek/Screenshot/Memory Alpha

**Star Trek's Holodeck:  
from science fiction to a  
new reality**

# Learning & assessment (L&A) has shifted to digital platforms

- **A lot as happened** in the **Generative AI** past few months.
- **Are we there yet?**
  - Digital-first learning & assessments (DfL&A)
  - Personalized & adaptive L&A
  - Game-based and VR/AR L&A => immersion
  - Collaborative L&A
- **What are novel fairness considerations for AI and AI-adjacent methods for digital L&A?**
  - Do we need novel fairness considerations for digital L&A with head-spinning momentum of innovation associated with large language models (e.g., GPT-4), and large multimodal models (e.g., image generation)?

# Generative AI



Large  
Computational  
Models  
(Language/Text  
Image/Voice/  
Video)

# ChatGPT swept us off our feet

**“Any sufficiently advanced technology is indistinguishable from magic”**

**— Arthur C. Clarke**

# Emerging Trends for Learning & Assessments

1. Holistic
2. Engaging, gamified, immersive
3. Personalized diagnostic & recommendations
4. Analytics for evaluation
5. Digital and multimodal delivery
6. Delightful experience (any time, anywhere)



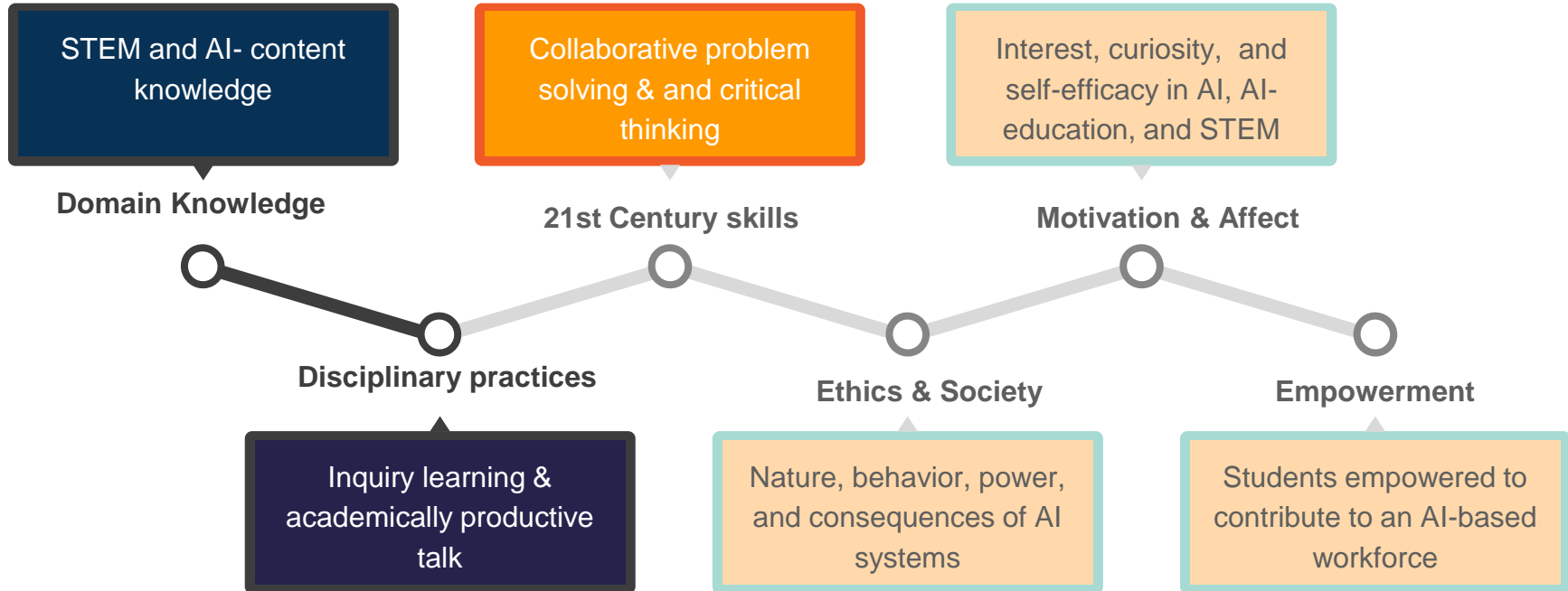
**The power is in the integration of frameworks & tools**



# iSAT integrates **AI-education** in science & tech courses to provide measurable learning outcomes\*

Peter Foltz & Sidney K. DMello  
Institute for Student-AI Teaming  
2023

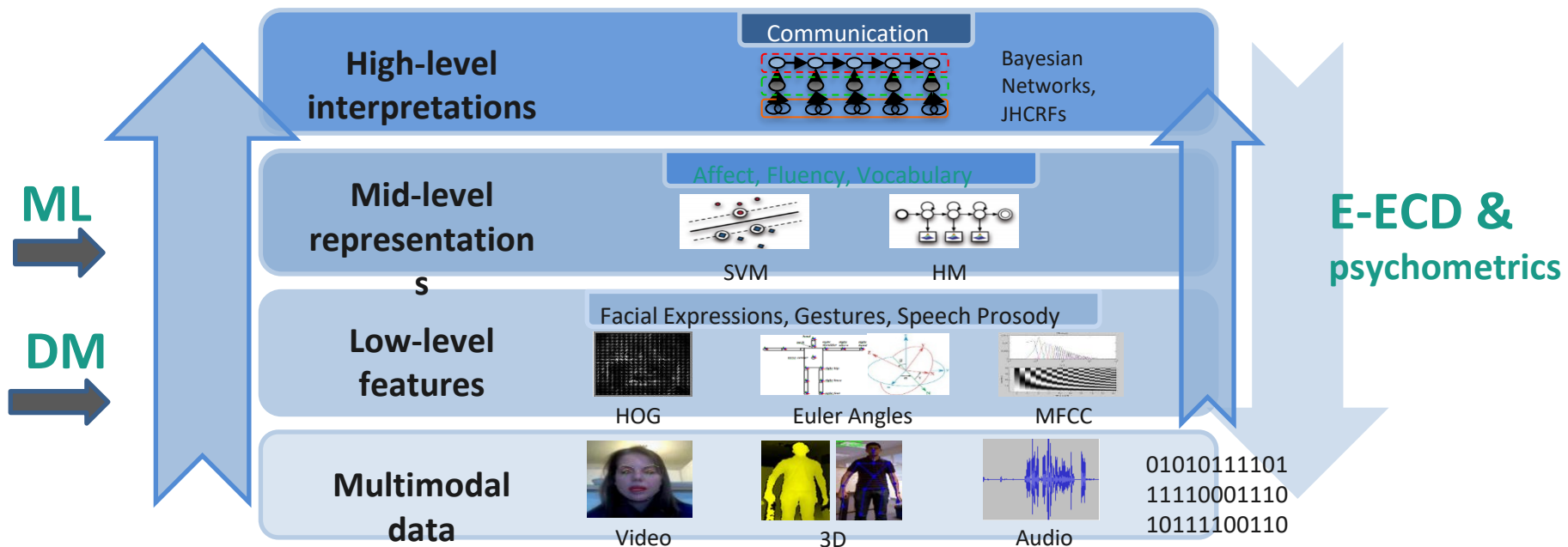
Institute for Student-AI Teaming  
University of Colorado Boulder  
[www.isat.ai](http://www.isat.ai)



\*Adapted slide

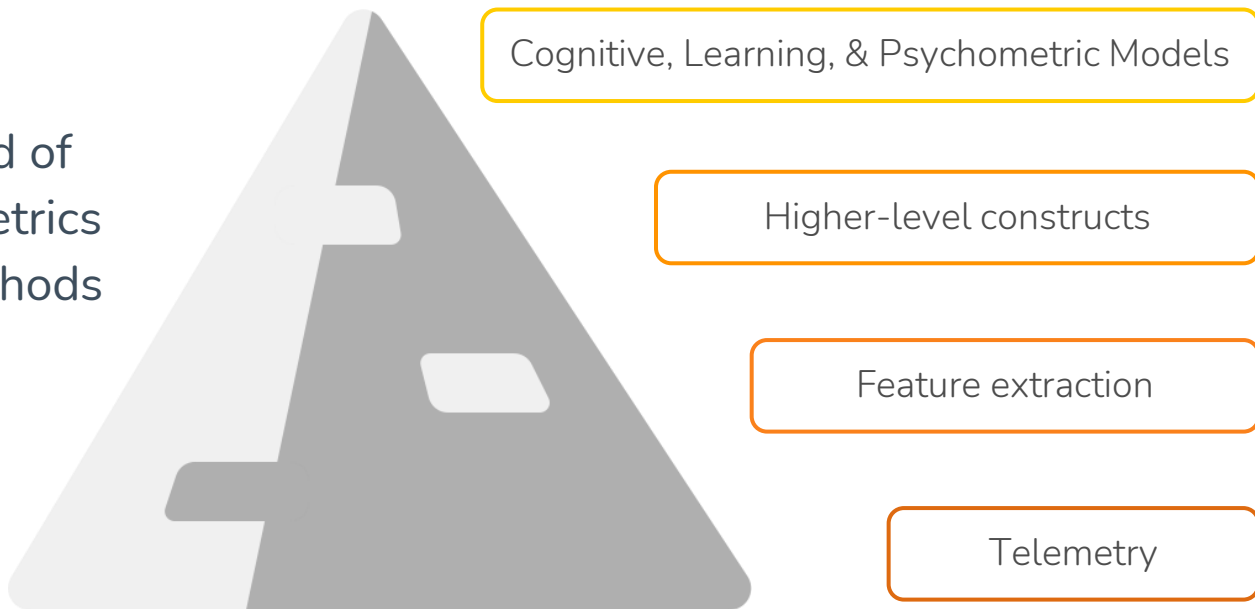
# Computational Psychometrics: Hierarchical Inference Model for Multimodal Data for L&A

- Design the system (learning &/or assessment) based on theory
- Identify **constructs** associated with competency of interest (*Expanded Evidence*)
- Find evidence for these constructs from low-level multiple sensory data (*Evidence about Evidence*) --Khan (2015); von Davier (2015)

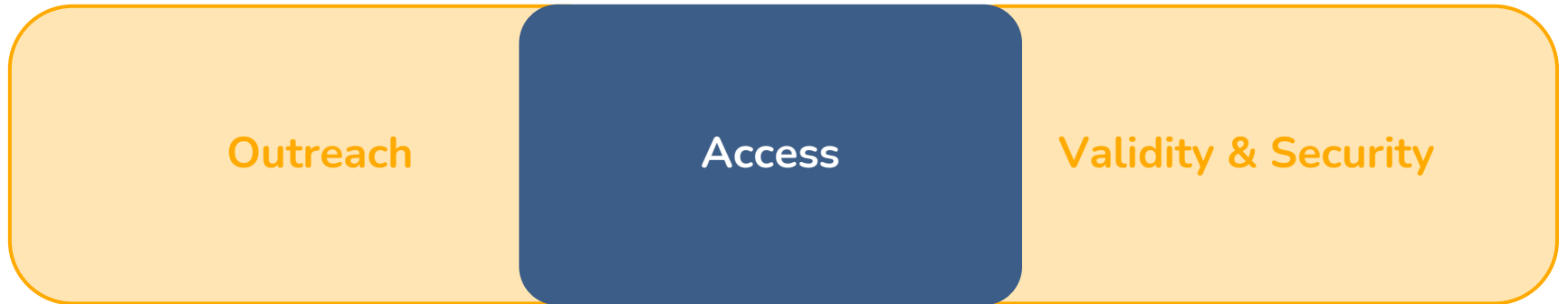


# Computational Psychometrics: An Integrative Framework for AI-based Capabilities

Computational psychometrics is a blend of theory-driven psychometrics & machine learning methods used to measure latent abilities in real time.



# We need a paradigm shift in assessment!



# Digital-first assessments as a solution for the new normal



“The future is already here. It’s just not evenly distributed, yet.”

— attributed to W. Gibson

# Characteristics of digital-first assessments

## Born digital

- Digitally-mediated constructs
- Designed with computational models & AI-algorithms in mind
- Theory-based & evidence-centered
- Valid & reliable

## Integrated

- Fluid infrastructure
- Integrative theoretical frameworks
- Integrated (automatic) tools & databases

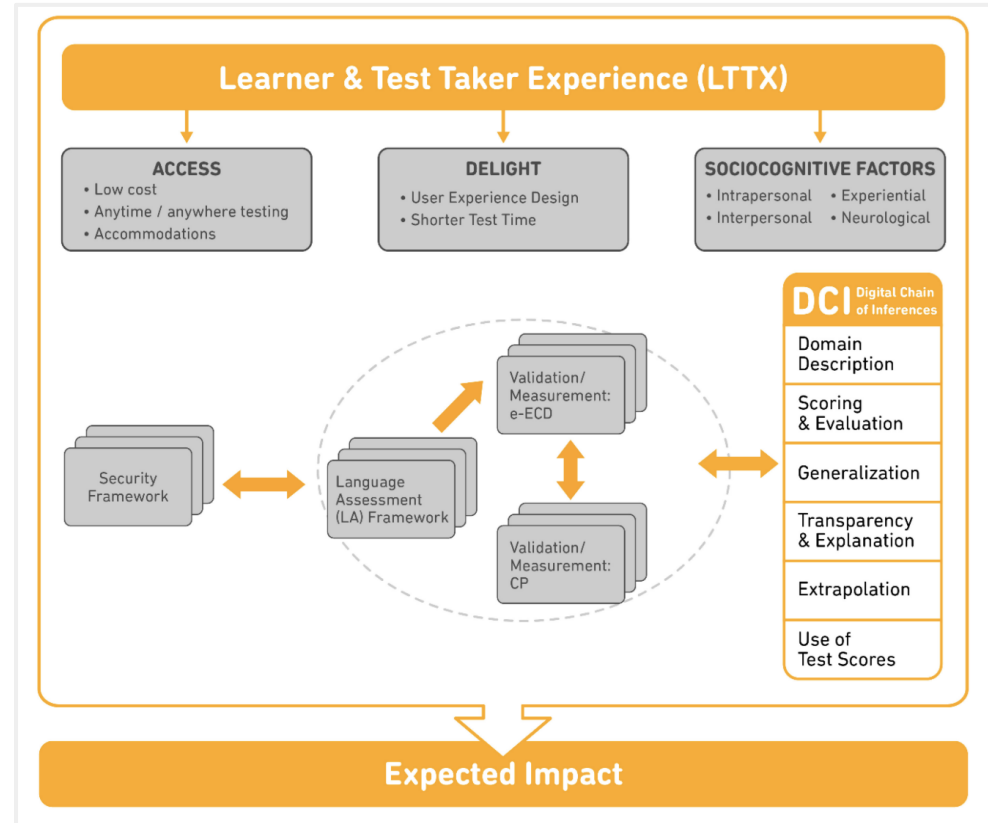
## Superior test-taker experience (TTX)

- Available on demand, anytime, anywhere
- Online testing with remote proctoring



# DLFA ecosystem

- Integrated learning & assessment frameworks
  - Digitally-mediated constructs
  - Security, Design, Measurement
  - LTTX
- Guiding questions for fairness auditing in the ecosystem
  - Where is AI used in the frameworks?
  - What practices support audits within & across frameworks?



# AI & ML within the Duolingo English Test

## Item development

- Content generation
- Question/prompt generation
- Option generation
- Review models (FAB & IQR)
- Review tools (Items factory & AQuAP)
- Pilot tools (Pre-pilot platform)

## Test administration

- Items microservice
- Adaptive algorithm
- Monitoring computer use
- Interactive tasks

## Scoring

- Automated speech recognition
- Feature development
- Scoring speaking
- Scoring writing
- Score monitoring (AQuAA)
- Pilot tools (Full-length practice test)

## Security

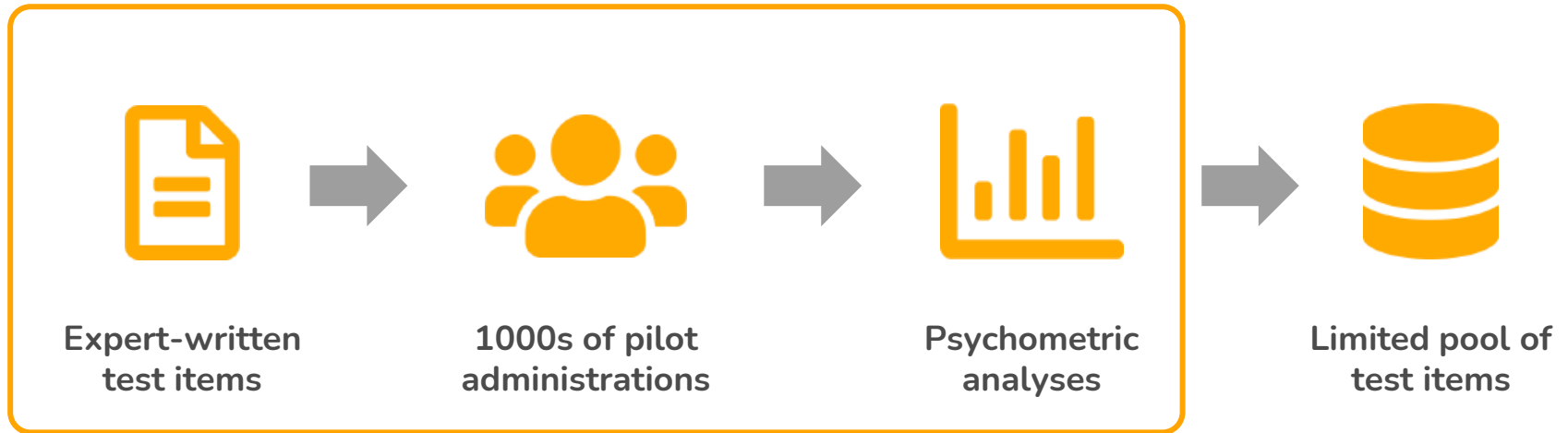
- Proctor portal
- AI behavior flags
- Computer use reports

## Score reporting

- Automated test taker report
- Score report API for test score users



# Traditional test development



Time-consuming & expensive

The pilot samples may not represent the target population

# Creation at scale: Human-in-the loop AI

## Goal: Humans leverage AI to perform a task

- Difficult for **both** humans and AI
- Have “expert” eyes on all content on the test

## “The Loop”:

- Humans design the items
- AI systems make predictions or generate content
- Humans review and/or edit the outputs
- Human labels and edits are used to improve AI

## At scale and with a high degree of efficiency

# Content creation at scale: AI-Driven Item Generation

Generative AI (GPT4, BERT, ChatGPT, Stable Difusion, etc)

can be used to automatically generate

- Fill-in-the-blank vocabulary questions
- Multiple choice keys and distractors
- Free response comprehension questions
- Writing and reading prompts
- Dialogues
- Images

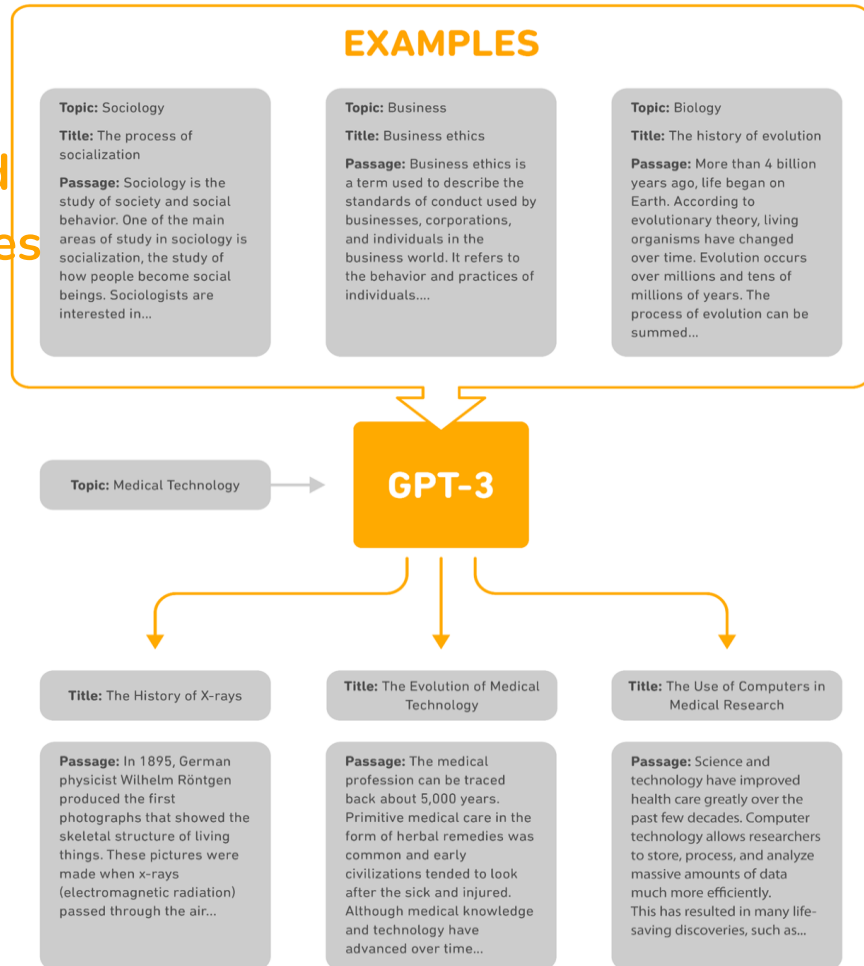
# Passage generation

GPT-3/4 can mirror style, format, and content given extremely few examples

## Goal

Generate self-contained passages in a range of domains - especially at an academic level

See Attali et al. (2022) for a detailed description of the creation of an interactive reading task



# Responsible AI for digital learning &

## assessment & Large Language Models

- Duolingo partnership with OpenAI
  - GPT-3 & GPT-4 for content generation
- **Responsible AI Standards for the Duolingo English Test (2023)**
  - First Responsible AI Standards for an assessment
    - Open for public comment
  - Aligned with DET's long term use of AI as a digital-first assessment

# The Future is Now!

Science fiction is any idea that occurs in the head and doesn't exist yet, but soon will, and will change everything for everybody, and nothing will ever be the same again.

— Ray Bradbury

DET 2023

# Thanks! Questions?

[avondavier@duolingo.com](mailto:avondavier@duolingo.com)

Acknowledgement:  
The Duolingo English Test R&D Area

