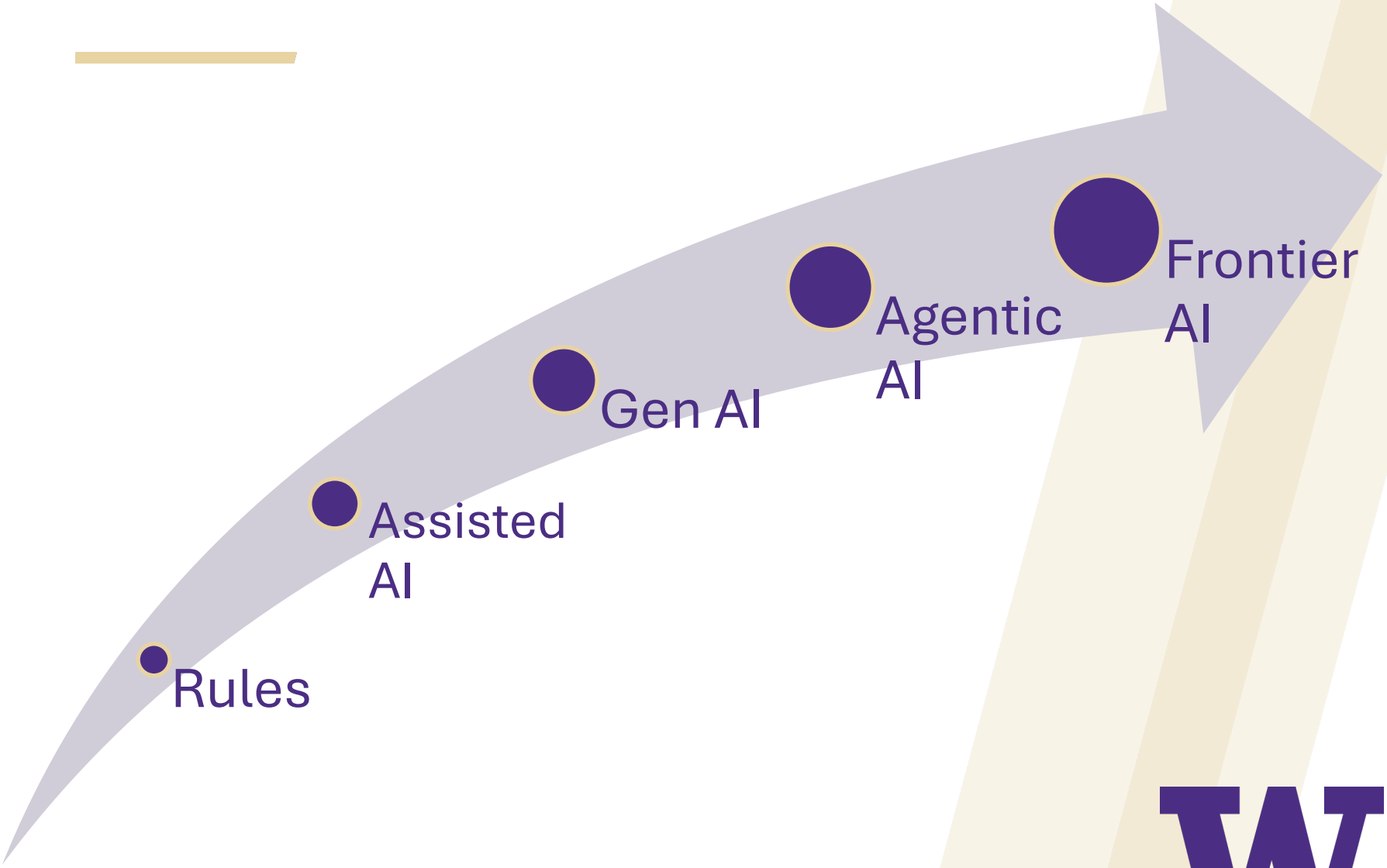
Beyond Black Boxes:
Case Studies of Transparent, Validated LLM Workflows for Accelerating Global HTA Submissions and Decisions

Beth Devine, Lockwood Taylor, Tim Reason, Bill Malcolm

ISPOR 2026 Philadelphia
Spotlight Session
Wednesday, May 20, 2026

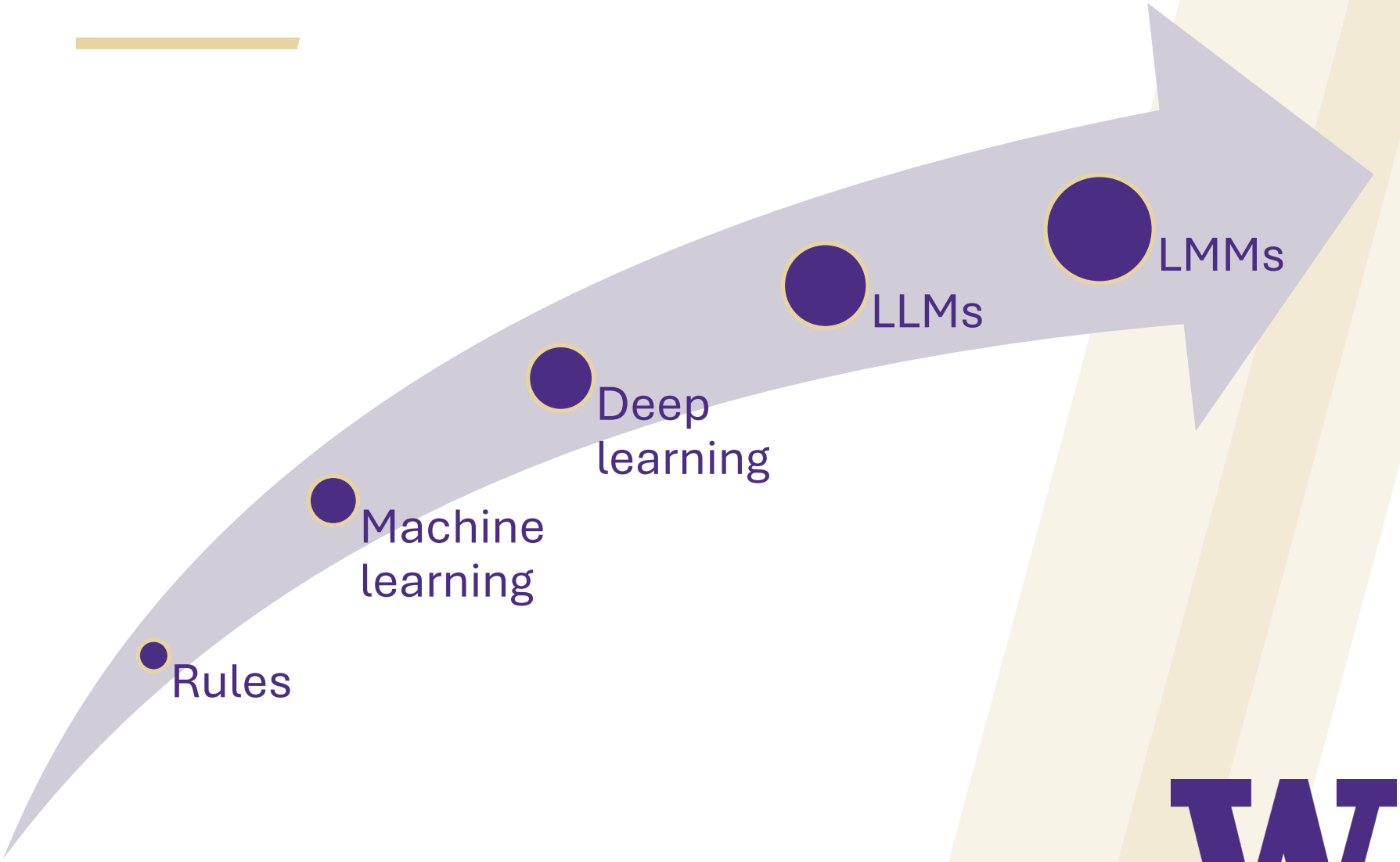
THE CHOICE INSTITUTE
School of Pharmacy

Evolution of AI: User interactions



Fleurence, et al. Taxonomy of Gen-AI in HEOR. *Value Health*. 2025; 28(11):1601–1610

Evolution of AI: From Rules to Foundation Models



Fleurence, et al. Taxonomy of Gen-AI in HEOR. *Value Health*. 2025; 28(11):1601–1610

RWE “Fit for purpose”

Relevant	• aligned with target trial question
Reliable	• accurate, complet traceabl
Robust	• calibrated and stab
Valid	• minimizes bias
Transparent	• understanda
Reproducible	• same results-sam
Replicable	• same results-diffe ent dataset

Equitable
Ethical
Private
Secure

FDA RWE Framework, 2018; FDA EHR/Claims RWD Guidance, 2024.
PMIDs: 38794897, 26994063, 33436424, 36215113, 26440803, 30429167, 38724471, 38871437.

Value in Health publications guide us (1)

ISPOR Report — VALUE HEALTH. 2024; 27(6):692–701

Assessing Real-World Data From Electronic Health Records for Health Technology Assessment: The SUITABILITY Checklist: A Good Practices Report of an ISPOR Task Force

Rachael L. Fleurence, PhD, MSc, Seamus Kent, PhD, Blythe Adamson, PhD, MPH, James Tcheng, MD, Ran Balicer, MD, PhD, MPH, Joseph S. Ross, MD, MHS, Kevin Haynes, PharmD, MSCE, Patrick Muller, PhD, Jon Campbell, MS, PhD, Elsa Bouée-Benhamiche, PharmD, RPh, Sebastián García Martí, MD, MSc, Scott Ramsey, MD, PhD

ISPOR Report — VALUE HEALTH. 2022; 25(7):1063–1080

Machine Learning Methods in Health Economics and Outcomes Research—The PALISADE Checklist: A Good Practices Report of an ISPOR Task Force

William V. Padula, PhD, Noemi Kreif, PhD, David J. Vanness, PhD, Blythe Adamson, PhD, Juan-David Rueda, MD, PhD, Federico Felizzi, PhD, MBA, Pall Jonsson, PhD, Maarten J. IJzerman, PhD, Atul Butte, MD, PhD, William Crown, PhD

Value in Health publications guide us (2)

ISPOR Report — VALUE HEALTH. 2025; 28(2):175–183

Generative Artificial Intelligence for Health Technology Assessment: Opportunities, Challenges, and Policy Considerations: An ISPOR Working Group Report

Rachael L. Fleurence, PhD, Jiang Bian, PhD, Xiaoyan Wang, PhD, Hua Xu, PhD, Dalia Dawoud, PhD, Mitchell Higashi, PhD, Jagpreet Chhatwal, PhD, on behalf of the ISPOR Working Group on Generative AI

ISPOR Report — VALUE HEALTH. 2025; 28(11):1601–1610

A Taxonomy of Generative Artificial Intelligence in Health Economics and Outcomes Research: An ISPOR Working Group Report

Rachael L. Fleurence, PhD, Xiaoyan Wang, PhD, Jiang Bian, PhD, Mitchell K. Higashi, PhD, Turgay Ayer, PhD, Hua Xu, PhD, Dalia Dawoud, PhD, Jagpreet Chhatwal, PhD, on behalf of the ISPOR Working Group on Generative AI

ISPOR Report — VALUE HEALTH. 2025; 28(11):1611–1625

ELEVATE-GenAI: Reporting Guidelines for the Use of Large Language Models in Health Economics and Outcomes Research: An ISPOR Working Group Report

Rachael L. Fleurence, PhD, Dalia Dawoud, PhD, Jiang Bian, PhD, Mitchell K. Higashi, PhD, Xiaoyan Wang, PhD, Hua Xu, PhD, Jagpreet Chhatwal, PhD, Turgay Ayer, PhD, on behalf of the ISPOR Working Group on Generative AI

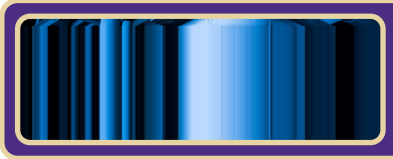
...and soon... Gen AI for SLRs

Uses and Applications

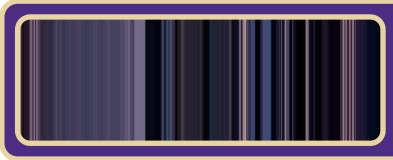
Evidence generation: clinical data abstraction

Evidence synthesis: network meta-analyses

Evidence synthesis: health economic modeling



Automated workflows/pipelines



The future???

Our spotlight today

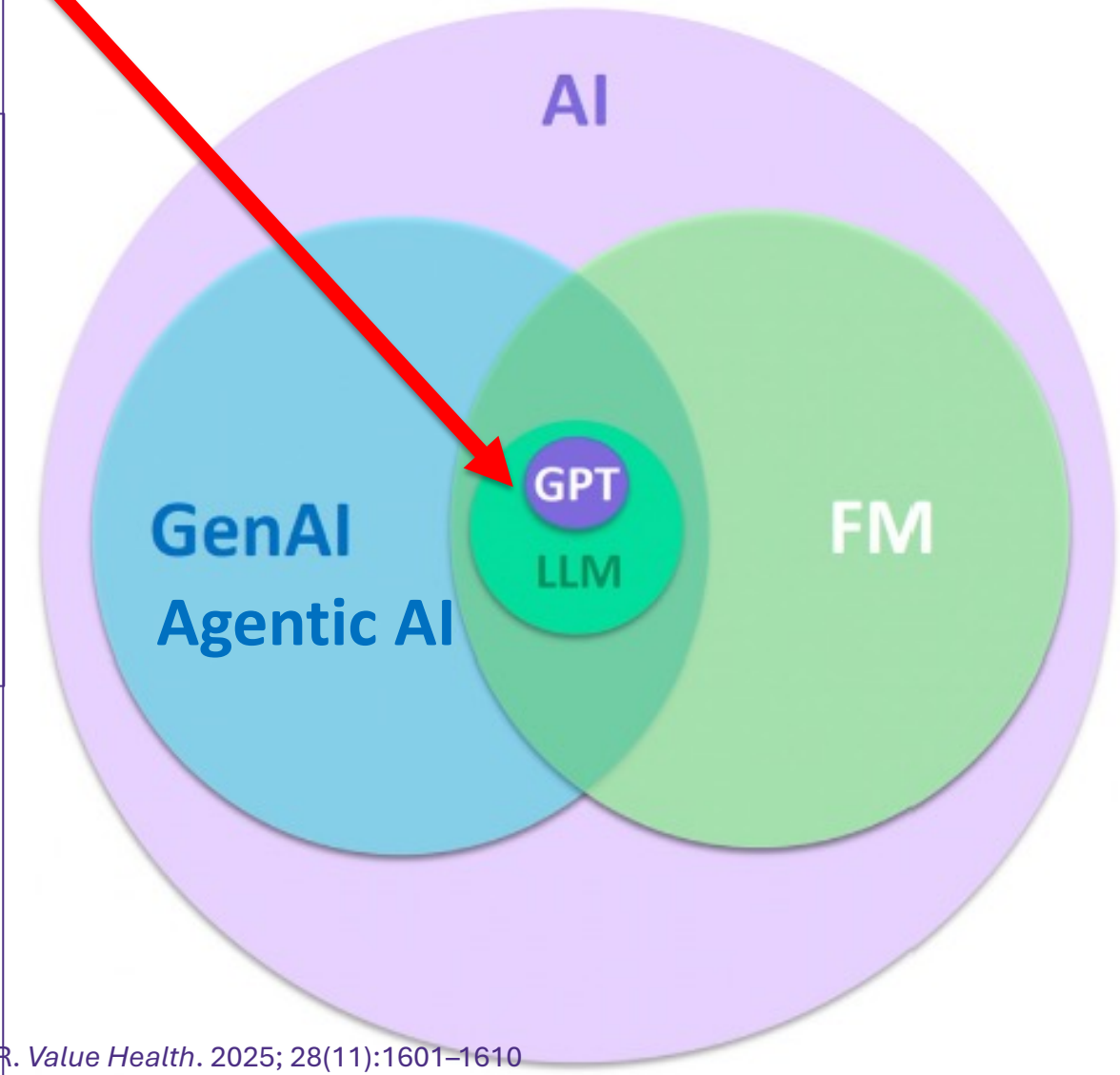
AI: artificial intelligence

FM: foundation model

Gen AI: generative AI

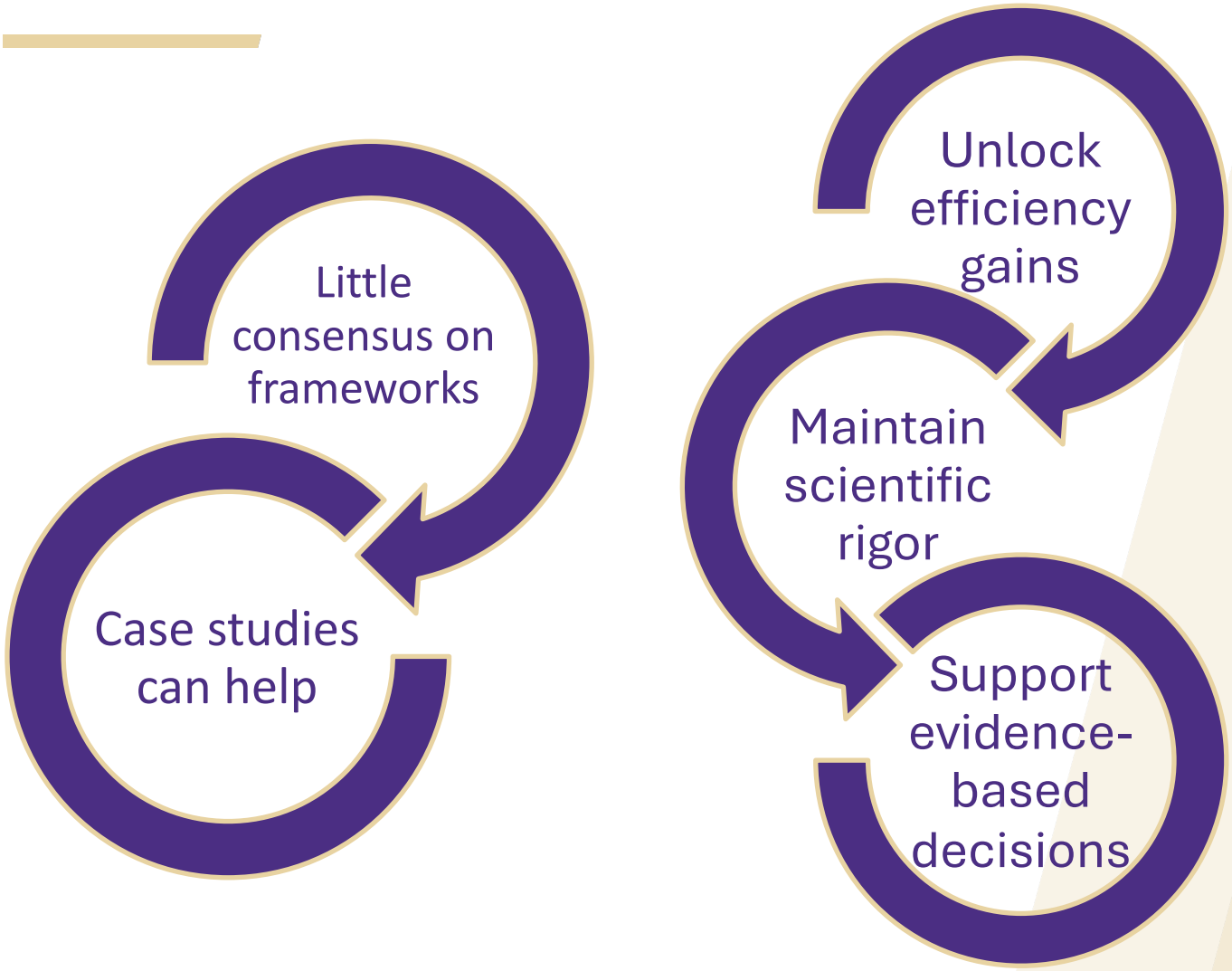
GPT: generative pre-trained transformer

LLM: large language model



Fleurence, et al. Taxonomy of Gen-AI in HEOR. *Value Health*. 2025; 28(11):1601–1610

Strengthening LLM frameworks



******Spotlight* Agenda! *****

- **Case study #1:** Application of LLMs to extract **systemic therapies**
  - **Unstructured** EHR data and physician insights
  - **VALID** framework
  - **multiple countries**

- **Case study #2:** Integration of validated LLM-generated data into AI workflows for **network meta-analyses**
  - highlighting transparency, reproducibility and HTA use

- **Case study #3:** Reduction of **HTA submission burden**
  - accelerating HTA submissions
  - maintaining quality standards
  - highlighting **LMIC contexts**

Your turn! Q & A/ Discussion

THE CHOICE INSTITUTE

School of Pharmacy

Thank you!

Beyond Black Boxes: Case Studies of Transparent, Validated LLM Workflows for Accelerating Global HTA Submissions and Decisions

Beth Devine, Lockwood Taylor, Tim Reason, Bill Malcolm

ISPOR 2026 Philadelphia
Spotlight Session
Wednesday, May 20, 2026

THE CHOICE INSTITUTE

School of Pharmacy

References (1)

Ensuring Reliability of Curated Electronic Health Record–Derived Data: The Validation of Accuracy for Large Language Model–/Machine Learning–Extracted Information and Data (VALID) Framework JCO Clin Cancer Inform 10:e2500215

Melissa Estevez, MS[1]; Nisha Singh, MS[1]; Lauren Dyson, MS[1]; Blythe Adamson, PhD, MPH[1]; Konstantin Krismer, PhD[1]; Kelly Magee, MS, FNP-BC[1]; Qianyu Yuan, PhD[1]; Megan W. Hildner, BSE[1]; Erin Fidyk, MSN, MBA[1]; Tori Williams, BA[1]; Olive Mbah, PhD, MHS[1]; Farhad Khan, MSPH[1]; Kathi Seidl-Rathkopf, PhD[2]; and Aaron B. Cohen, MD, MSCE[1]

DOI https://doi.org/10.1200/CCI-25-00215

PharmacoEconomics - Open (2024) 8:205–220
https://doi.org/10.1007/s41669-024-00476-9

ORIGINAL RESEARCH ARTICLE

Artificial Intelligence to Automate Network Meta-Analyses: Four Case Studies to Evaluate the Potential Application of Large Language Models

Tim Reason[1] · Emma Benbow[1] · Julia Langham[1] · Andy Gimblett[1] · Sven L. Klijn[2] · Bill Malcolm[3]

PharmacoEconomics - Open (2024) 8:191–203
https://doi.org/10.1007/s41669-024-00477-8

ORIGINAL RESEARCH ARTICLE

Artificial Intelligence to Automate Health Economic Modelling: A Case Study to Evaluate the Potential Application of Large Language Models

Tim Reason[1] · William Rawlinson[1] · Julia Langham[1] · Andy Gimblett[1] · Bill Malcolm[2] · Sven Klijn[3]

Pharmaceutical Medicine (2024) 38:365–372
https://doi.org/10.1007/s40290-024-00539-6

ORIGINAL RESEARCH ARTICLE

Automated Mass Extraction of Over 680,000 PICOs from Clinical Study Abstracts Using Generative AI: A Proof-of-Concept Study

Tim Reason[1] · Julia Langham[1] · Andy Gimblett[1]

References (2)

1. US FDA. Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products: Guidance for Industry. Silver Spring (MD): US FDA; 2024.
2. US FDA. Framework for FDA's Real-World Evidence Program. Silver Spring (MD): US FDA; 2018.
3. Wang, et al. . Building transparency and reproducibility into the practice of pharmacoepidemiology and outcomes research. Am J Epidemiol. 2024;193:1625-1631.
4. Hernán, et al. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol. 2016;183:758-764.
5. Wang, et al. STaRT-RWE: structured template for planning and reporting on the implementation of real-world evidence studies. BMJ. 2021;372:m4856.

References (3)

6. Wang,et al. HARmonized Protocol Template to Enhance Reproducibility of Hypothesis Evaluating Real-World Evidence Studies on Treatment Effects: A Good Practices Report of a Joint ISPE/ISPOR Task Force. Value Health. 2022;25:1663-1672.
7. Benchimol, et al.; RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015;12(10):e1001885.
8. Langan, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). BMJ. 2018;363:k3532.
19, Gini, et al. Describing diversity of real-world data sources in pharmacoepidemiologic studies: the DIVERSE scoping review. Pharmacoepidemiol Drug Saf. 2024;33(5):e5787.
9. Fleurence, et al. Assessing real-world data from electronic health records for health technology assessment: the SUITABILITY checklist: a good practices report of an ISPOR task force. Value in Health. 2024 Jun 1;27:692-701. PMID: 38871437.