U.S. FOOD & DRUG ADMINISTRATION

Regulatory Plenary

AI-Based Tools in Psychiatry Drug Development

A regulatory framework for clinical trial implementation

American Society of Clinical Psychopharmacology | Miami

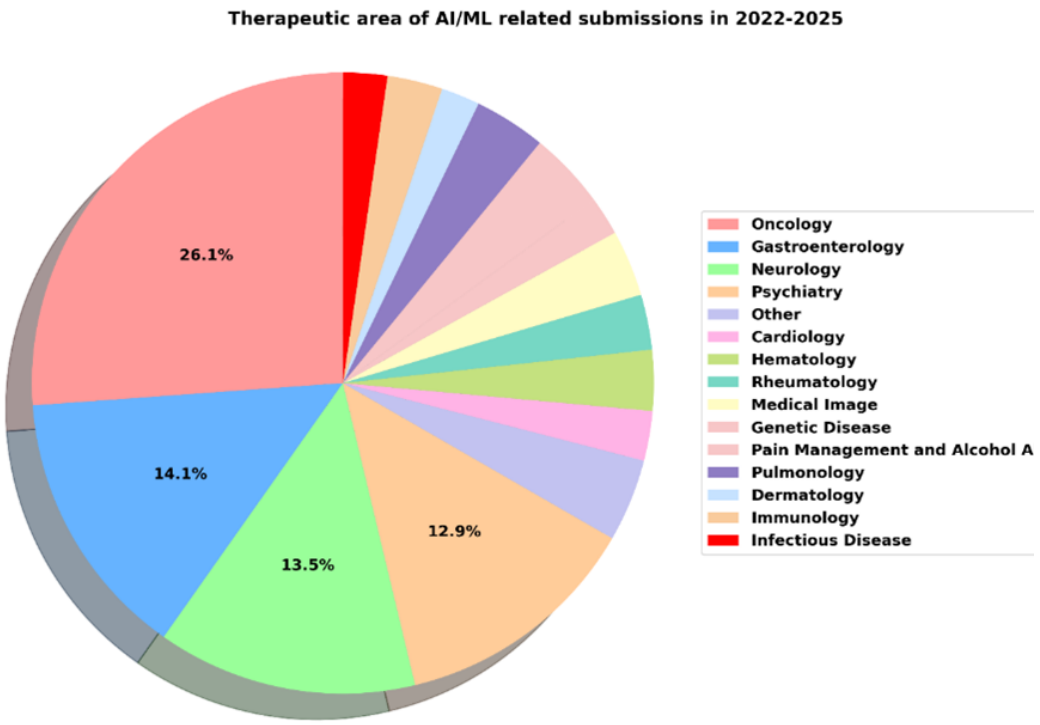
Valentina Mantua, MD, PhD
Associate Director for Regulatory Science
Office of Neuroscience, CDER

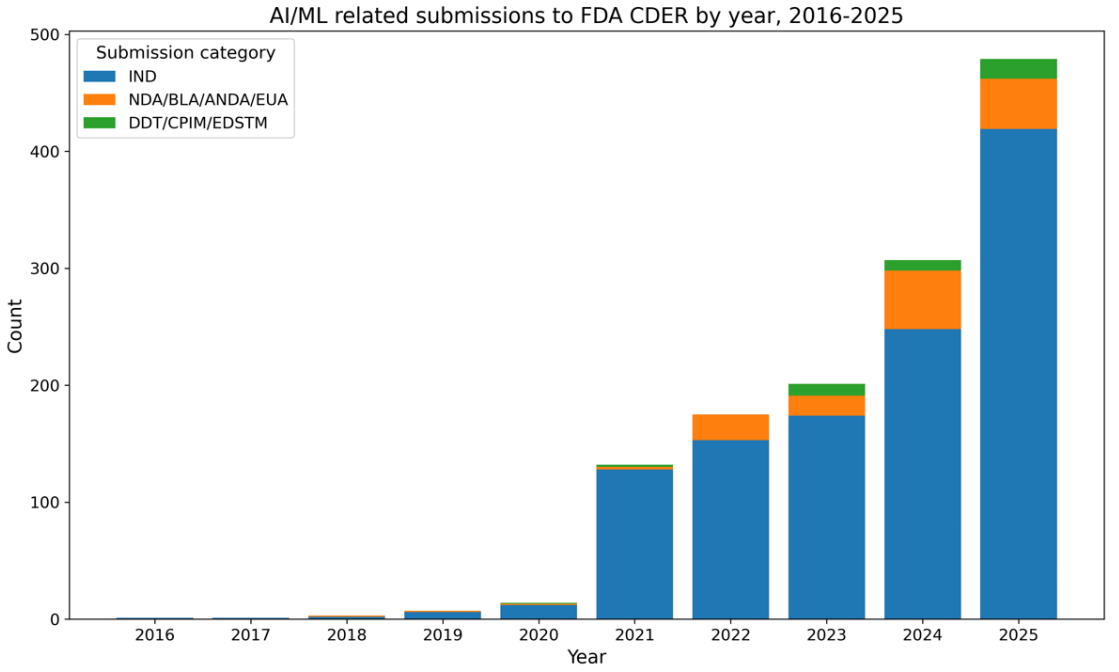
Disclaimer: The views expressed are my own and should not be interpreted as official FDA policy.

Landscape Analysis for AI/ML-Related Submissions at CDER

Landscape analysis includes submissions in years 2022-2025

AI/ML submissions are increasing and broadening across objectives and data modalities.





Why this matters for neuroscience:

AI is moving from exploratory analytics toward trial design, measurement, and evidence generation. The regulatory question becomes: What role does the model play, and how much confidence is needed for that role?

Figure credit: Dr. Menglun Wang. Landscape analysis builds on Liu et al., Clin Pharmacol Ther. 2023.

AI in Drug Development: Broad Uses, Neuroscience Focus

AI is part of the drug product life cycle — Potential uses across the drug product life cycle.

Nonclinical	Clinical	Postmarketing	Manufacturing/Inspections
Reducing the number of animal-based pharmacokinetic, pharmacodynamic, and toxicologic studies	Using predictive modeling for exposure-response analyses Processing and analyzing large sets of data for the development of clinical trial endpoints or assessment of outcomes	Identifying, evaluating, and processing for reporting post marketing adverse drug experience information Adverse-event triage/RWD analytics	Discovery Enhancing quality control Generating data to document and manage processes subject to regulatory oversight and inspection

Narrowing to neuroscience clinical trials:

1 Learn biology

Multiscale models connect biology → brain → behavior: omics, fluid biomarkers, imaging, DHTs, clinical endpoints.

2 Model trajectories

Subtypes, staging, progression, placebo response and uncertainty become trial-ready hypotheses.

3 Design precision trials

Stratify/enrich patients, simulate outcomes, optimize endpoints, and adapt the Phase 2 → Phase 3 path.

Sources: FDA, Artificial Intelligence for Drug Development; FDA AI regulatory decision-making guidance.

Known Challenges of Drug Development in Psychiatry

Limited disease understanding. Unlike many medical fields, psychiatry often lacks clear neurobiological targets, making it difficult to design mechanism-based therapies

Diagnostic ambiguity. The boundaries between psychiatric diagnoses remain blurred, and current nosographic frameworks may not reflect the underlying biology, complicating trial design and endpoint selection.

Measurement gap — Higher-order brain functions — thought, mood, energy, volition, perception — are extraordinarily difficult to quantify reliably, leaving clinical trials dependent on subjective, often inconsistent rating tools.

We Have the Technology. Can we trust it?

Unprecedented technological capability. AI now offers remarkable reasoning power and the computational capacity to integrate complex, multi-source data at a scale previously unimaginable in psychiatric research.

New tools for behavioral measurement. Longitudinal, continuous, and objective measurement of behavior is now within reach, addressing one of psychiatry's most persistent methodological limitations.

The central challenge: trustworthiness. As AI capabilities outpace our familiarity with them, the critical question shifts from availability to validity, reliability, and fitness for purpose in high-stakes clinical trial contexts.

FDA's role This talk focuses on how FDA views AI-based outputs in the context of drug development clinical trials, and how the agency has developed an evolving, risk-proportionate regulatory framework to evaluate and qualify these technologies responsibly.

Source: Presenter framing; FDA AI guidance considered final by presenter.

Considerations for AI Use in Regulatory Decision-Making

FDA's risk-based credibility framework.

1 **Question of interest**
What decision, question, or concern will the AI output inform?

2 **Context of use**
What will be modeled, and how will the output be used?

3 **Model risk**
Model influence + decision consequence.

4 **Credibility plan**
Evidence needed to establish model credibility within the COU.

5-7 Execute the plan, document the results and assess deviations, determine adequacy..

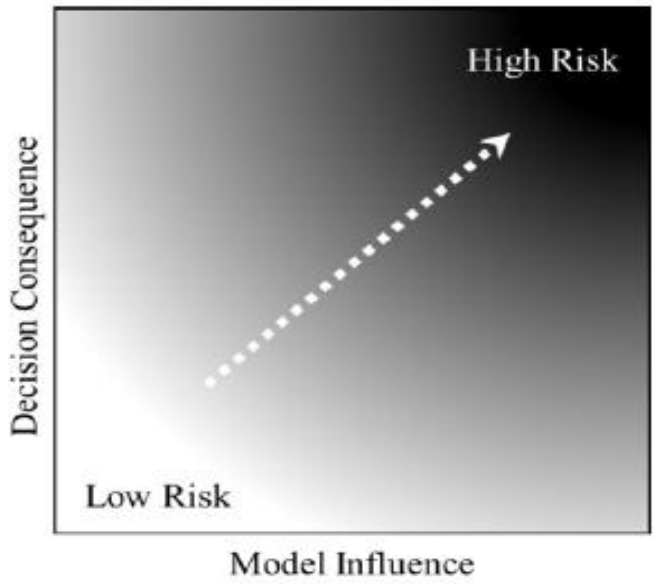


Figure 1. Model risk matrix. The model risk moves from low to high as decision consequence or model influence increases. The ratings for decision consequence and model influence are independently determined.

Model influence
How much the AI-derived evidence contributes relative to other evidence used to inform the question of interest.

Decision consequence
Significance of an adverse outcome from an incorrect decision concerning the question of interest

Credibility assessment activities should be commensurate with model risk and tailored to the context of use.
The risk-based credibility assessment framework envisions interactive feedback from FDA concerning the assessment of the AI model risk (step 3) as well as the adequacy of the credibility assessment plan (step 4) based on the model risk and the COU.

Source: FDA Draft Guidance on the Use of Artificial Intelligence To Support Regulatory Decision-Making for drugs and biologics

Determine the Adequacy of the Credibility Assessment

Does the credibility assessment plan support the AI model output for a specific context of use?

If credibility is not sufficiently established for the model risk, the pathway is iterative:

Downgrade influence
Use AI output with other evidence; add human oversight; narrow its role.

Increase rigor
Add fit-for-use data; strengthen independent testing; quantify uncertainty.

Add controls
Pre-specify thresholds; monitor performance; mitigate drift and failure modes.

Revise or reject COU
Change the model, endpoint, or analysis; keep using exploratory if necessary.

What FDA expects to see
A credibility assessment report that connects: question of interest -> context of use -> model risk -> planned assessment -> results and deviations -> adequacy conclusion.

No single method is prescriptive; the evidence must support the intended use.

Source: FDA Draft Guidance, Step 7.

Anonymized Examples: Model Influence Shapes FDA Feedback

Examples are illustrative. The same AI method can require different evidence depending on its role in the trial.

Example A: LLM-derived scoring used for rating an outcome scale	Example B: ML-optimized DHT features
Proposal LLM-derived scoring will serve as a key secondary endpoint *Consensus Rater* score to complement and support the primary endpoint of human individual *Site Raters.*	**Proposal** Use in-clinic and at-home digital health technology features to train an ML-derived score summarizing change over time for use as an exploratory endpoint.
Model Influence High. No human oversight is proposed in using LLM-derived score. It is the sole factor to inform the question of interest.	**Model Influence** High. The output is a composite score that is planned to serve as an effectiveness endpoint in clinical trials with the target patient population and is the sole factor to inform the question of interest.
Decision Consequences (If LLM-derived score makes inaccurate decision): Low in Phase 2, as it is not intended to support regulatory decision-making. Moderate in Phase 3.	**Decision Consequences** Low to High depending on COU. Exploratory characterization is different from an endpoint supporting effectiveness. As the COU expands, model influence and decision consequence can increase.
Feedback The sponsor agreed to use the LLM only as an exploratory endpoint in the current program and work to strengthen the credibility assessment. The sponsor plans to use the LLM-derived score as the primary endpoint for a future study in a different program.	**Feedback** Before it can be used as an endpoint: clinically meaningful ground truth, fit-for-use data, independent test set, subgroup and longitudinal performance, and a prespecified analysis plan.

Regulatory feedback is most productive when the sponsor frames the question of interest, COU, model risk, and credibility evidence up front.

Sources: FDA AI regulatory decision-making framework; anonymized examples adapted from slide materials provided by presenter.

Participant's baseline data can be operationalized as a baseline prognostic score used for covariate adjustment — start with the regulatory use case, not the algorithm.

Question of interest

Can a baseline prognostic score improve the precision of treatment-effect estimation in a randomized trial?

Context of use

Model trained on independent historical data, fixed before analysis, and used as a prespecified covariate.

Phase 2: exploratory. Phase 3: potentially primary.

Model risk

Influence: high. Consequence: low.

In randomized trials, weak scores may yield no gain — or some loss — in precision.

Proposed credibility plan

Representative training/validation data and independent evaluation.

Locked model, prespecified SAP, repeatable score generation, and explainability on key drivers.

Illustrative FDA feedback

- No objection for Phase 2 exploratory use.
- Phase 3 primary analysis may be acceptable with full trial and SAP details.
- Show that the training data represent the population to be enrolled.
- If generative AI is used, ensure the same patient retains the same value if the process is repeated multiple times.

Sources: https://www.fda.gov/media/148910/download; anonymized sponsor feedback provided by presenter.

Engagement Pathways

Bring four elements early: question of interest, context of use, model risk, and a proposed credibility plan.

Use case	Potential engagement option
Specific development program	IND / pre-IND formal meeting with review division
AI in trial design or statistics, RWE	IND or other formal meetings, C3TI, Complex Innovative Trial Design, MIDD
AI-enabled DHT, endpoint, or DDT	IND or other formal meetings, CPIM, ISTAND
General methodology / technology	CPIM, CDER AI Council

Critical Path Innovation Meetings (CPIM)
CDER Center for Clinical Trial Innovation (C3TI)
Innovative Science and Technology Approaches for New Drugs (ISTAND)
Model-Informed Drug Development Paired Meeting Program (MIDD)

Talk To US!

Sources: FDA AI regulatory decision-making guidance Appendix 2; CDER Center for Clinical Trial Innovation (C3TI) webpage.

Do We Use AI at FDA? Yes, but human subject matter experts decide.

FDA uses AI to support scientific, operational, and surveillance work - while regulatory decisions remain human decisions.

Artificial Intelligence / ML
Adverse-event data mining
Predictive and computational toxicology
Review analytics and prioritization

Generative AI
Elsa enterprise platform
Summarization and label comparisons
Code and database support

Agentic AI
Optional multi-step workflow support
Human oversight built in
Secure deployment environment

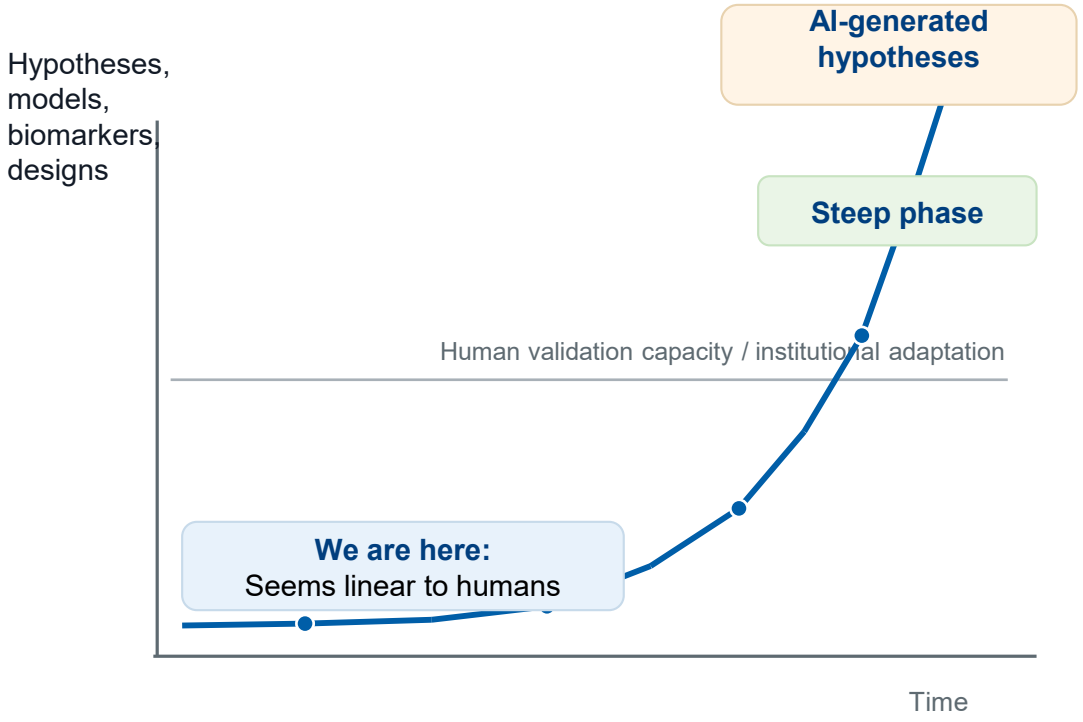
AI supports. Subject matter expert decide
- AI outputs are assistive.
- Outputs must be validated by staff with relevant expertise and domain knowledge.
- Regulatory decisions or actions are made by human experts.
- Data safeguards and security controls matter.

The same principle applies externally: AI can accelerate work, but credibility and accountability remain essential.

Sources: FDA press announcements on Elsa (June 2025) and agentic AI deployment (Dec. 2025); FDA LEADER 3D training material; HHS AI use case inventory.

The Exponential Challenge for Regulatory Science

Today, AI can still be handled like another development tool. The harder possibility is that increasingly capable systems generate and refine candidate scientific hypotheses faster than we can validate them.



Today, regulatory science is largely evidence review: assessing submitted evidence against standards for safety, effectiveness and quality.

The nature of evidence changes: more AI-generated hypotheses and simulations - the bottleneck shifts to validation.

Regulatory science becomes evidence governance: preserve the conditions for trustworthy evidence: validation, traceability, transparency, and accountability.

The regulator of the future will be the guardian of trustworthy evidence (?)

Conceptual closing perspective; AI can generate candidate hypotheses, but validation remains the limiting step.

U.S. FOOD & DRUG ADMINISTRATION

Thank you

Questions / discussion

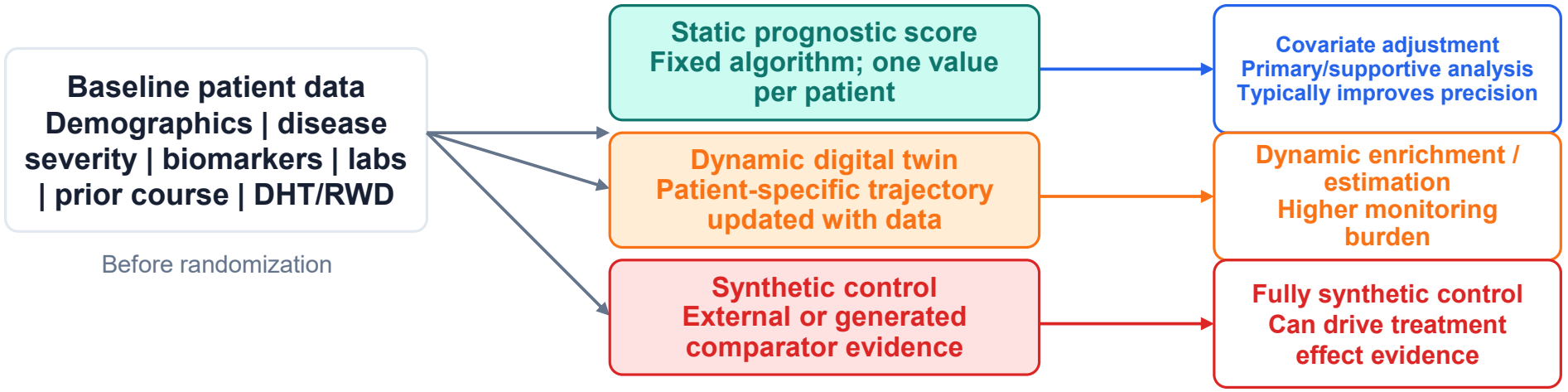
Valentina.Mantua@fda.hhs.gov

Disclosure: OpenAI ChatGPT assisted with slide creation; content reviewed and revised by presenter.

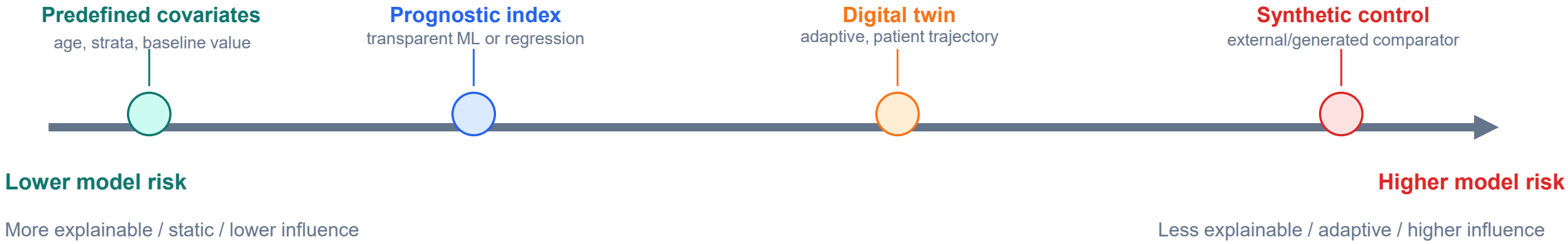
U.S. FOOD & DRUG ADMINISTRATION

Back up

AI-derived covariates in randomized trials: role, transparency, and regulatory risk



Spectrum: lower risk when explainable, prespecified, and used only to improve precision; higher risk when opaque, adaptive, or highly influential



Baseline data can support covariate adjustment, but model credibility expectations should scale with context of use.