

# Juan D. Pinto

juan@jdpinto.com | jdpinto.com | github.com/juandpinto

## EXPERIENCE

---

### McGraw Hill

Remote

*Data Scientist*

*Aug 2025 – Present*

- Built student-behavior segmentation system using latent profile analysis on 200K+ students and thousands of courses, achieving 93% classification confidence and enabling mid-semester classification by week 4 via FIML
- Led causal impact study of 2K+ students using GLMMs and propensity-weighted difference-in-differences, finding that early, high AI-feature adoption significantly increased weekly engagement
- Drove LLM/RAG evaluation across 9 judge LLMs and 4 embedding models on a 500-question suite (15,421 retrievals), delivering a 3-LLM jury aligned with SME labels ( $\kappa=0.737$ ) and selecting the top retriever
- Spearheaded cross-functional R&D for instructor-facing analytics and robust learner proficiency models

### AI Institutes Virtual Organization (AIVO) + Google.org

Remote

*Google.org AI4Ed Research Fellow*

*May 2025 – Aug 2025*

- Designed multi-agent generative AI storytelling platform for screening specific learning disabilities (SLDs)
- Created modular agentic framework that includes robust guardrails, custom states, and adaptive agent delegation

### NSF AI Institute for Inclusive Intelligent Technologies for Education

Urbana, IL

*Learner Modeling Graduate Research Assistant*

*Aug 2023 – Aug 2025*

- Developed predictive models of student skills and behaviors for real-time adaptive learning
- Led data cleaning and analysis on datasets with tens- to hundreds-of-thousands of student actions

### Human-Centered Educational Data Science Lab (HEDS)

Urbana, IL

*Graduate Research Assistant*

*Sept 2020 – May 2025*

- Developed predictive and inferential models of student behaviors, emphasizing explainable AI
- Investigated CS students' coding patterns using epistemic network analysis, LLMs, and various ML approaches
- Authored 13 peer-reviewed publications (6 as lead author) in venues related to *educational data science*

### ETS Research Institute

Princeton, NJ

*Ida Lawrence Research Intern*

*June 2024 – July 2024*

- Developed small heuristic classification models (for ensembling) that detect student reading disengagement
- Validated models indirectly (unlabeled data) using response accuracy, on-task behavior, and book preferences

## PROJECTS

---

### Interpretable Neural Network for Learner Behavior Detection | *Python, PyTorch*

- Developed a convolutional neural network for detecting rare gaming-the-system behavior among learners
- Emphasized interpretable-by-design approach via custom loss function and novel thresholding mechanism
- Demonstrated that the model provides fully faithful explanations utilizing 100% of its inference-time parameters and 90% explanation intelligibility among human users

### Weight-Based Modeling for Student Performance Prediction | *Python, Scikit-learn, PyTorch, TensorFlow*

- Engineered complex weighted features to predict student performance on future coding problems
- Demonstrated that logistic regression with weighting schemes matched SOTA model performance
- Won 2nd place in the 2022 Educational Data Mining in CS Data Challenge

## EDUCATION

---

### University of Illinois Urbana-Champaign

Urbana, IL

*Ph.D. in Educational Data Science*

*Aug 2025*

### University of Michigan

Ann Arbor, MI

*M.A. in Design and Technologies for Learning*

*July 2020*

### University of Texas at Austin

Austin, TX

*M.A. in Middle Eastern Languages and Cultures (Language Acquisition)*

*May 2018*

### Brigham Young University

Provo, UT

*B.A. in Ancient Near Eastern Studies (minor in Linguistics)*

*May 2016*

## SKILLS

---

**Data Analysis:** Python (NumPy, Pandas, PySpark, Polars, Matplotlib, Seaborn), R, SQL

**Machine Learning & AI:** Scikit-learn, PyTorch, TensorFlow, Keras, LangChain, LangGraph

**Cloud & Data Platforms:** AWS, Databricks, BigQuery