# Juan D. Pinto

801-228-7184 | juan@jdpinto.com | jdpinto.com

## Education

**University of Illinois Urbana-Champaign** — Urbana, IL
*Ph.D. in Educational Data Science* — *Aug 2020 – Jul 2025*

**University of Michigan** — Ann Arbor, MI
*M.A. in Design and Technologies for Learning* — *Aug 2019 – July 2020*

## Experience

**Learner Modeling Graduate Research Assistant** — Aug 2023 – Present
*NSF AI Institute for Inclusive Intelligent Technologies for Education (INVITE)* — *Urbana, IL*
- Developed predictive models of student non-cognitive skills and behaviors
- Conducted cleaning and analyses on large-scale educational datasets from varied digital platforms

**Graduate Research Assistant** — Sep 2020 – Present
*Human-Centered Educational Data Science Lab (HEDS)* — *Urbana, IL*
- Led and participated in various research projects in the fields of *learning analytics* and *educational data mining*
- Developed predictive and inferential models of student behaviors, with an emphasis on explainability

**Ida Lawrence Research Intern** — Jun 2024 – Jul 2024
*ETS Research Institute* — *Princeton, NJ*
- Developed simple interpretable models (for ensembling) that detect student disengagement in a reading app
- Validated models using accuracy, on-task behavior, and book preferences

## Projects

**Interpretable Neural Network for Learner Behavior Detection** | *Python, PyTorch*
- Developed a convolutional neural network for detecting gaming-the-system behavior in education
- Emphasized interpretable-by-design approach via loss-term regularization and novel thresholding mechanism
- Demonstrated that the model provides fully interpretable explanations faithful to its learned knowledge
- Evaluated the model's performance and explainability against human expert-identified patterns

**Evaluating LLMs for Debugging Strategy Classification** | *Python*
- Developed pipeline for systematic LLM prompting across different dimensions, such as *chain-of-thought*, *zero-* vs. *few-shot*, *single-* vs. *multi-label*, *reasoning*, and *fine-tuned*
- Outperformed baseline methods in classifying students' debugging strategies
- Improved annotation efficiency and minority class detection in student code

**Weight-Based Modeling for Student Performance Prediction** | *Python, Scikit-learn, PyTorch, TensorFlow*
- Developed weighting schemes to predict student performance using programming traces
- Designed similarity metrics based on code, problem prompts, and struggling patterns
- Showed that *source code* and *struggling pattern* similarity, along with *problem order*, improved prediction accuracy
- Demonstrated that logistic regression with weighting schemes matched deep-learning performance

**Modeling Student Performance Using Measures of Persistence** | *Python, Scikit-learn*
- Developed a random forest model to predict student quiz performance
- Conducted careful feature engineering based on previously studied elements of student persistence
- Analyzed features and their interactions using SHAP values

**Epistemic Network Analysis of CS Students' Debugging Behavior** | *R*
- Analyzed debugging behaviors in novice programmers using Epistemic Network Analysis (ENA)
- Identified key constituents of the debugging process based on expert interpretation of student behaviors
- Compared debugging strategies between students with different prior programming experience
- Investigated how debugging behaviors evolved over time as students gained experience in a CS1 course

## Technical Skills

**Data Analysis & Visualization**: Python (NumPy, Pandas, Matplotlib, Seaborn), R, SQL
**Machine Learning & AI**: Scikit-learn, PyTorch, Tensorflow, Keras
**Web Programming**: HTML, CSS, Javascript, Hugo, Jekyll
**Other**: Git, LaTeX, Adobe Creative Cloud (Photoshop, Illustrator)