

## Accelerating Pace of SUD Research

- **Student-Centered Impact:** PSGV students are experienced professionals that transform raw data into actionable insights daily
- **Educational Imperative:** Leverage Artificial Intelligence (AI), Machine Learning (ML) and Data Analytics related coursework to embed Substance Use Disorder (SUD) research problems.
- **Societal Challenge:** With over 100 000 annual opioid-related deaths in the U.S., mobilizing student talent accelerates SUD research innovation.

## Graduate Professionals Engagement

- **SUD Data Integration:** Ingest, clean, and harmonize multi-source SUD datasets as part of C<sup>ross</sup> Industry Standard Process for Data Mining (CRISP-DM)
- **Modeling & Analytics Proficiency:** Implement CRISP-DM–based workflows, from exploratory analysis to predictive modeling, with an emphasis on bias mitigation.
- **Collaborative Innovation:** Foster interdisciplinary teamwork, pairing data science, AI, ML, public health, and clinical experts, to co-develop recommendations systems and policy dashboards.
- **Ethical & Privacy Awareness:** Instill best practices for HIPAA/42 CFR Part 2 compliance, data de-identification, and responsible AI in SUD research.

## Adapted CRISP-DM in the Classroom

- **Phase 1 - Business & Research Understanding:** Teams define success metrics like e.g.: treatment adherence rates, relapse reduction percentages.
- **Phase 2 - Data Exploration & Preprocessing:** Tasks like e.g.: de-identify free-text notes, engineer socio-demographic features, visualize dispensing trends.
- **Phase 3 - Modeling & Recommender Prototyping:** Build hybrid recommenders combining content-based, collaborative, and Large Language Models (LLM) - driven suggestions.
- **Phase 4 - Evaluation & Iteration:** Conduct A/B testing on historical SUD data; interpret performance and equity metrics.
- **Phase 5 - Deployment & Reflection:** Develop dashboards and APIs showcased in lab demos; teams report on lessons learned and future improvements.

## Data Pipeline Design and Student Roles

- **Flowchart TB**  
A [Data Ingestion] --> B [Student-Curated Data Lake]  
B --> C [Streaming ETL (Apache Flink)]  
C --> D [Embedding Service (BioClinicalBERT)]  
D --> E [Feature Store]  
E --> F [Recommender Engine]  
F --> G [Student-Facing Dashboard]  
G --> H [Feedback Loop]  
H --> C
- **Possible Student Roles:**
  - Project Managers: Oversee the project
  - Software Developers: Build modules
  - Data Engineers: Build and monitor ingestion streams
  - NLP Analysts: fine-tune LLMs for clinical context
  - Visualization Designers: craft interactive dashboards

## Measurable Student Impact

- **Innovation Diffusion:** Through the courses and Nittany AI Challenge, 10+ teams created SUD research-focused prototypes, geospatial overdose heatmaps, SMS-based relapse alerts, and harm-reduction chatbots
- **Coursework Integration:** 30 student teams across three years performed descriptive, predictive and prescriptive data analysis
- **Research Assistant Contributions:** Graduate RAs built the backbone of igntd.com SPARx framework, implemented scalable Flink streaming pipelines, reducing ETL latency by 70% and enabling near-real-time insights
- **Skill Growth:** 85% of participants report increased proficiency in Data Analytics, AI, Python, SQL, and ML frameworks.

## Possible Collaborations and Scaling

- **Expanded Partnerships:** Collaborate with CSUA affiliates to validate recommender outputs in live settings.
- **Interdisciplinary Courses:** CSUA teams directed prototyped dashboard views, geo-analytics, and early recommender demos
- **Mentorship Networks:** Establish CSUA mentoring for sustained knowledge transfer in SUD research.

**Collaborators include:** Dusan Ramljak, Nonna Sorokina, Abhay Haridas, Tanuja Voruganti, Faizan Muhammad Raza, Jainil Kakka, Tayo Obasanmi, Ahmed Soliman, Mark Fahim, Hunter Berberich, Brandon Botzer, Justin Bright, Kevin Durkin, Allison Gray, Chenese Gray, Kethrie Heasley, Lesley Krewatch, Ananthakrishnan Krishnan, Meng Li, Alcides Licona, Zining Lyu, Tyler Mankey, Arielle Markiewicz, Ani Rao, Deependra Singh Rawat, Victoria Ritorto, Christopher Sahr, Sean Smith, Vladimir Sveshnikov, Matthew Ulrich, Ray Zaremski, Pat Tento, Rudraksh Mishra, Nikhitha Kunduru, Ambika Chundru, Shaheed Mahbub, Vishwas Murali, Sricharan Boduppali, Gayatri Bangar, Falguni Rathore, Bharat Sharma, Danica Chanda, Tiithi Agrawal, Gaurang Prabhudesai, Alex Brulliea, Tyler Staffin, Labib Nazer, Alvin Mathew, James Grissel, Anna Forrest, Savitha Kolar, Seyed Pouya Mahdavi, Jeff Oxenberg, and Satish Srinivasan– Penn State Great Valley, Pierce Baugh and Adi Jaffe – igntd.com