PIER Prospectus: Trustworthy AI to Optimize Cyber-Physical Systems and Enhance Social Justice

1. Executive Summary

College/School Name: Roy G. Perry College of Engineering

Lead Dean: Dr. Pamela Obiomon

Proposed Research Theme: Advanced Technologies and Artificial Intelligence (AI)

Summary of Initiative:

Brief Description of the Interdisciplinary Focus

By uniting the expertise of the College of Engineering, the College of Arts and Sciences, and the College of Juvenile Justice, we would like to launch an interdisciplinary initiative on Trustworthy Artificial Intelligence (AI). Together, these colleges will ensure that AI systems powering cyber-physical systems, such as smart grids, autonomous vehicles, medical devices, and robotics, are safe, reliable, secure, and fair. By combining technical innovation with ethical insight and justice perspectives, the team will tackle challenges like protecting critical infrastructure from cyberattacks, reducing bias in automated decision-making, and improving transparency in criminal justice. This initiative strengthens PVAMU's role as a national leader in AI research, expands opportunities for students, and delivers real-world impact in energy, transportation, healthcare, and social equity.

Artificial Intelligence (AI) and machine learning have emerged as transformative forces in our lives and they have the potential to revolutionize various aspects of our society, from cyber-physical systems (CPS) such as smart grids, robotics, and autonomous vehicles to criminal justice. To harness the power of trustworthy AI for industrial revolution and societal well-being, an interdisciplinary team of researchers from three colleges (Engineering, Arts & Sciences, Juvenile Justice) at Prairie View A&M University (PVAMU) propose a project on ***Trustworthy AI to Optimize Cyber-Physical Systems and Enhance Social Justice.*** *The overarching goal is to perform research and education in trustworthy AI and integrate the technologies with domain knowledge to address the challenges in CPS and social justice.* This ambitious interdisciplinary initiative will integrate engineering and computer science with social and behavioral sciences to *advance foundational principles of trustworthy AI* as well as domain knowledge in CPS and juvenile justice when developing the solutions, leveraging the complementary expertise of the team. Together the proposed project will address the urgent needs in CPS and social justice using a holistic solution enabled by the emerging technologies and tools of trustworthy AI. *The developed methodologies and tools can be extended to many applications that of interests to all the colleges of PVAMU such as in precision agriculture and healthcare*. The project will leverage CREDIT center's strengths in AI, machine learning, big data analysis, and edge computing as well as the research centers such as the SECURE and SMART centers to develop AI systems that are transparent, secure, reliable, and equitable.

Key Societal Challenges Addressed:

By aligning technical integrity and robustness with ethical principles, trustworthy AI bridges the gap between technological advancement and societal well-being, ensuring that the benefits of AI will optimize many CPS such as critical infrastructure, and will be distributed fairly across all segments of society. Specifically, the project will address the challenges in

- Reliability and robustness of AI
- Explainability of AI
- AI fairness and accountability in automated decision-making
- Cybersecurity and resilience of AI applications in critical infrastructures (e.g., energy, transportation)
- Ethical and inclusive use of AI in justice systems and workforce development

Anticipated Impact:

The initiative will elevate PVAMU's national visibility in AI research by fostering interdisciplinary collaboration, positioning PVAMU for federal and philanthropic research investment, and training a diverse next-generation AI workforce through postdoc and student engagement. It integrates teaching, research, and outreach through real-world AI deployment in smart infrastructure and justice systems, and encourages broader participation in AI revolution from the entire PVAMU community.

Alignment with Journey to Eminence:

This initiative supports Strategic Priorities 2 (Teaching Excellence), 3 (Student Success), and 4 (Research Growth) by addressing urgent national needs in AI revolution while cultivating innovation, ethics, and interdisciplinary excellence. It also aligns well with the America's AI Action Plan.

2. Alignment with Strategic Plan

Strategic Priorities Supported:

This project, *"Trustworthy AI to Optimize Cyber-Physical Systems and Enhance Social Justice,"* aligns closely with Prairie View A&M University's Strategic Plan, specifically supporting the following strategic priorities:

- Priority 2: Advance Teaching Excellence and Academic Relevance
- Priority 3: Enrich Student Success and Holistic Development
- Priority 4: Drive Strategic Advancement of Research This project

This project is expected to advance teaching excellence through the integration of emerging technologies into the curriculum (Priority 2), enrich student success through hands-on research and mentorship (Priority 3), and strengthen the university's research enterprise in high-impact, socially relevant areas (Priority 4). By addressing technological innovation and integrate research into teaching, the project reflects PVAMU's commitment to academic relevance, student development, and solutions that benefit both local and global communities. This project also aligns very well with the *America's AI Action Plan*: https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf

Intercollege Linkages:

This initiative demonstrates strong intercollege linkages through interdisciplinary collaboration within the **Roy G. Perry College of Engineering (COE)** and across multiple colleges, including the **College of Arts and Sciences (CAS)** and the **College of Juvenile Justice (CJJ)**. While drawing on expertise in AI, machine learning, big data analysis, and cyber-physical systems (CPS) such as smart grids, robotics, and autonomous vehicles in engineering (COE), faculty from CAS will contribute perspectives on trustworthy AI through the lenses of psychology and cognitive science, while researchers from CJJ will address issues of fairness, justice, and the societal impacts of AI. Together, this collaborative effort aims to provide a holistic solution to pressing challenges in trustworthy AI research and education.

Synergies & Leverage:

This initiative leverages complementary expertise of researchers from multiple colleges to advance beyond current AI practices. In this project, foundational knowledge will be gained in trustworthy AI in terms of reliability, robustness, explainability, and security. In addition, novel methods to prevent, detect, and mitigate harmful biases in AI systems will be designed and implemented using a socio-technical approach. This endeavor will foster a deep understanding of both the capabilities and limitations of emerging AI

technologies. Moreover, it will provide indispensable guidance for the design and development of Generative AI tools and Large Language Models (LLMs) that align with scientific principles and human and societal values, made possible through cross-disciplinary collaboration. The outcomes will advance AI research and education, enhance public trust in AI applications, and benefit society as a whole.

The Roy G. Perry College of Engineering proudly hosts twelve outstanding research centers, each contributing significantly to advancing knowledge and innovation in various domains such as artificial intelligence, cybersecurity, the smart grid, climate change, and computational biology. For example, the CREDIT center, established in 2015 with $6 million in seed funding from the Department of Defense, focus on mission-critical big data analytics and AI research. Over the years, the center has actively collaborated with government agencies, academia, and industry partners, contributing significantly to both research and education. Multiple laboratories have been established under its leadership. Building on the established capabilities and facilities of the CREDIT center and other research centers, this initiative, *Trustworthy AI to Optimize Cyber-Physical Systems and Enhance Social Justice,* will further expand research and education capacity through collaborations with College of Arts and Sciences (CAS) and the College of Juvenile Justice (CJJ), fostering interdisciplinary engagement and innovation.

The synergy among the proposed research and educational objectives and activities is illustrated in Figure 1.

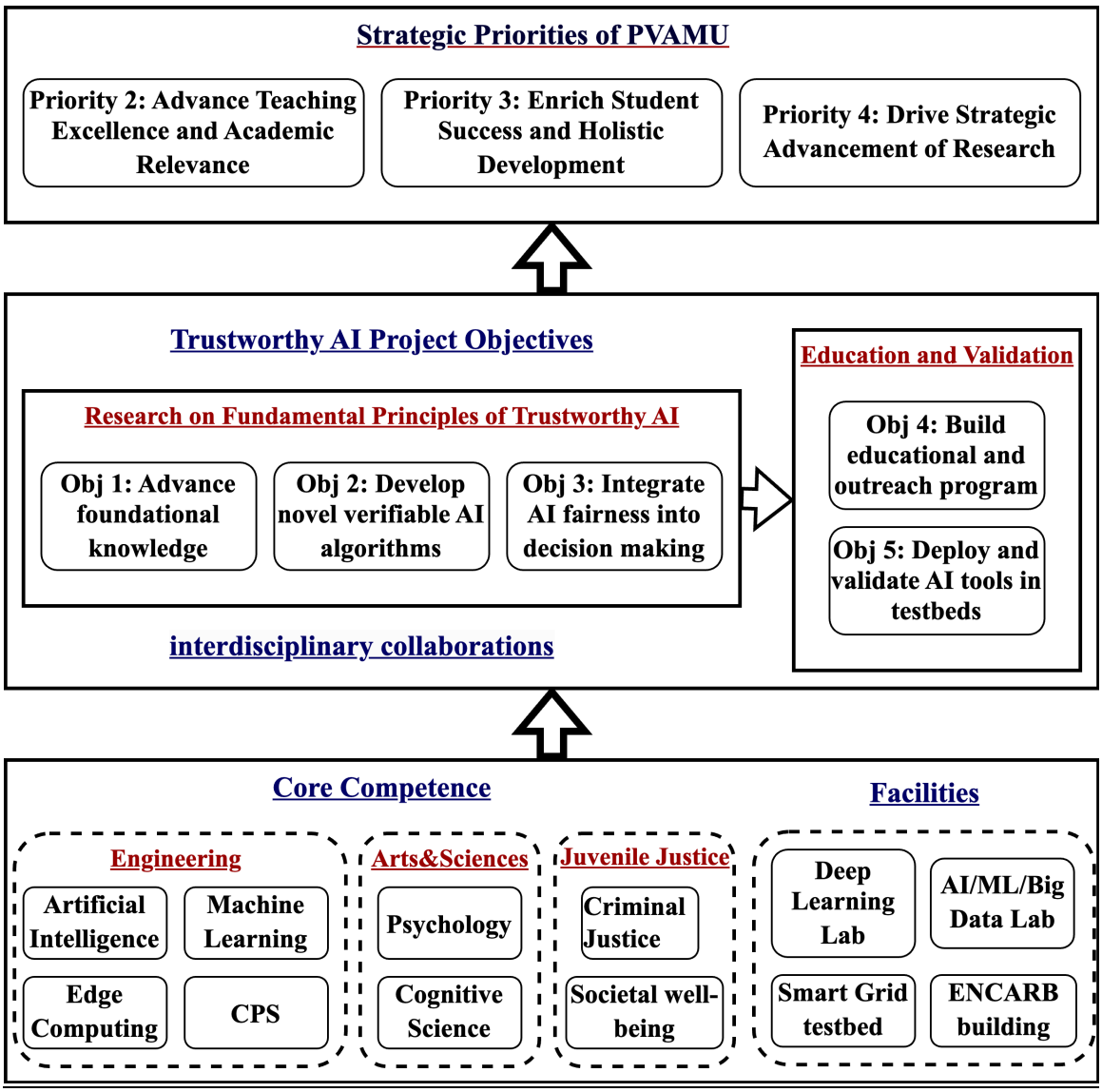

Figure 1. Synergy among the proposed research and education objectives and activities.

3. Research Vision and Goals

Big-Picture Vision:

To lead nationally and internationally recognized trustworthy AI research and education, especially in the advancement of foundational knowledge of trustworthy AI and in the development and deployment of trustworthy AI systems that solve real-world challenges in critical infrastructure and ensure fairness in AI applications across social systems.

Specific Objectives:

1. Advance foundational knowledge of trustworthy AI (lead: Engineering):

- Reliability and Robustness
- Explainability
- Resilience and Security/Privacy
- Fairness, Bias Mitigation, and Societal Well-being

Reliability and Robustness

Reliability and robustness of emerging AI systems are of outmost importance for mission-critical applications. Although there have been a lot of studies in this topic, there exist many challenges to be solved: (i) Out-of-distribution generalization; (ii) Robustness to noisy data; (iii) Robustness in transfer learning. In this project, we plan to (i) investigate techniques to ensure AI models generalize well to unseen, out-of-distribution data. (ii) Study preprocessing techniques to clean and enhance data quality, and (iii) propose novel methods to avoid catastrophic forgetting when transferring knowledge across different domains.

Explainability

Powerful, complex machine learning (ML) models, such as deep neural network models, are usually involved with millions or billions of parameters and difficult to interpret. With the wide adoption of neural networks and other ML models in mission-critical applications, explainable AI/ML models are not only important but also necessary for better understanding and decision-making. In this project, we will explore four approaches for interpreting neural networks: (i) Explaining individual feature importance by computing input gradients or by using post-hoc means, (ii) Developing attention-based models, which illustrate where neural networks focus during inference, (iii) Visualization of feature map, de-convolution, and saliency maps; and (iv) Indirect interpretation that mimics the prediction of neural networks using simpler and interpretable models.

Resilience and Security/Privacy

In order to provide security and resilience to attacks in trustworthy AI, the team plans to explore the following research topics: (i) Adversarial training: two types of adversarial training will be considered, 1) Projected gradient descent adversarial training, and 2) Trading-off accuracy and robustness via distributionally robust optimization; (ii) Computational resource-limited federated learning. Many mission critical applications involves training and testing ML models directly on edge devices, where computational power is highly restricted due to hardware and power limits. Thus, the design of federated learning algorithms must satisfy the constraint of computational resources on edge devices and alleviate on-edge computation costs. We plan to explore several solutions to address the computational resource limit on edge devices, including but not limited to layer disentanglement, model pruning, model distillation, and progressive learning. The proposed method will be examined for its privacy protection properties and compared with federated learning with differential privacy.

Fairness, Bias Mitigation, and Societal Well-being

While Generative AI has undergone substantial progress, it can mirror and perpetuate societal biases, encompassing stereotypes and prejudices. These models lack a contextual understanding comparable to

human comprehension, potentially resulting in biased or inappropriate responses. In this project, we plan to address architecture bias, emergent bias, and AI-generated content bias introduced by Generative AI and model selection bias during algorithm design. Specifically, (i) Mitigating architecture bias through explainable AI via conversational model of explanation: This task aims to enhance the explainability of GAI architectures by incorporating Monarch Mixer and human-in-the-loop. More importantly, this task intends to implement innovative prompt tuning through a conversational model of explanation; (ii) Mitigating emergent bias through integrating fairness metrics into learning objective: Emergent bias occurs when an AI system is utilized beyond its initially planned domain of application. We propose to integrate novel fairness metrics into the learning functions to mitigate emergent bias. The learning functions will incorporate fairness metrics, such as the disparity impact from IBM AIF360 as regularization terms, where these fairness metrics will be defined through collaboration between social scientists and engineers/computer scientists. (iii) Mitigating AI-generated content bias via collaborative AI: Generative AI models may generate content that is offensive, inappropriate, or reflective of societal biases. This challenge is particularly pronounced due to the limited human involvement in current methods for detecting AI-generated content. The objective of this task is to pioneer a novel collaborative AI methodology that seamlessly integrates the cognitive capabilities of both humans and AI agents to detect biases in AI-generated content. Additionally, these AI agents can autonomously update model parameters in a swarm learning fashion, without central control.

2. Develop verifiable AI algorithms for smart grid and autonomous systems (lead: Engineering)

Developing verifiable AI algorithms for Cyber-Physical Systems (CPS) is critical considering the safety and security requirements of those systems. In this project, we plan to adopt the hybrid systems model, where both continuous dynamics and discrete switching are integrated to model CPS. In addition to using the traditional reachability analysis, a novel reservoir computing approach is proposed to create a digital twin of the CPS. An Echo State Network model will be trained using both simulated data and real-world data.

3. Integrate ethical frameworks and AI fairness tools into decision models (lead: Arts & Science, Juvenile Justice)

The process of integrating ethical frameworks and AI fairness tools into decision models involves four stages. The pre-design stage will focus on planning, problem specification, and identification of data. It requires translating high-level objectives or strategic goals into tractable problems, necessitating the identification of appropriate target variables and datasets. These choices are rarely self-evident, normative assessments often take them for granted, even though different translations can raise profoundly different ethical concerns. Since available data may differ significantly from what occurs in the real world due to sampling bias, historical bias, and systemic bias, the availability vs. representativeness tradeoff will be studied. In the design stage, the use of datasets in AI applications must be adapted to take into the full spectrum of socio-technical factors of the context in which they are deployed into consideration. In this task, we propose to study diverse sampling techniques such as stratified random sampling for Internet-scale data used by LLMs to ensure that each group is adequately represented, as well as new methods for effectively generating stratified partitions of datasets. For problems with very limited data, data augmentation techniques will be investigated for creating fair synthetic data and increasing the amount of data for the sensitive groups using adversarial learning such as FairGAN, thus expanding the diversity of

the dataset, and substantiating surrogate data while taking into account known limitations. In the development stage, we plan to study techniques such as suppression, dataset massaging, and reweighing samples. However, excluding protected attributes can often be correlated with proxy attributes that remain in the dataset where bias may still be present. Hence, we plan to study how to select data to optimize the underlying construct rather than optimize based on the proxies. In addition, we propose to address the decontextualizing data using socio-technical analysis to provide insights into social variations in the dynamics of a phenomenon. Challenges and novel solutions of trustworthy AI in the deployment stage will also be investigated with human and society feedback.

4. Build educational and outreach programs focused on trustworthy AI (All)

The objectives of the proposed educational and outreach activities include: (i) Curriculum development: provide strong support of the new Master of Science in Data Science and Engineering (MSDSE) program at PVAMU, and create new courses in trustworthy AI such as a new graduate course titled "Generative AI and Foundation Models"; (ii) Lab enhancement: enhance the new AI/ML/Big Data Lab in the ENCARB building by developing simulations and experiments for students to conduct their lab activities associated with this project and related courses; (iii) Students' mentoring: the faculty will closely advise students on their academic plans and career development. For instance, they will assist students with manuscript preparation, improving presentation skills, and provide information about summer internship. Students will be encouraged to attend seminars and workshops regularly, as well as participate in professional meetings, to interact with other researchers and build their professional network; (iv) Seminar series: a seminar series on trustworthy AI will be organized where top researchers in the field will be invited to deliver talks and faculty, postdoc, and students are encouraged to present their research results and exchange ideas; (v) Training workshops: the team will collaborate with the industrial partners such as IBM and NVIDIA to offer training tutorial workshops such as the Deep Learning and Generative AI workshop to outreach to broad audience and train future AI workforce.

5. Deploy and evaluate AI prototypes in CPS and justice-related applications with real-world data (All)

In this project, the teams from all colleges will work together to (i) deploy pretrained robust machine learning models in the smart grid testbed to predict electricity demand and execute reinforcement learning enabled optimization to maximize efficiency and the usage of renewable energy sources while minimizing the failure rate; (ii) collect real-world datasets for social systems such as the juvenile justice system, apply the proposed bias mitigation methods and interpretable machine learning models to validate and demonstrate the fairness, transparency, and accountability of trustworthy AI in a reimagined system.

Innovation through Convergence:

This project bridges AI design, ethics, and technical strengths in edge computing and generative AI, combined with PVAMU's community connections, allow for rapid, ethical prototyping with measurable societal impact.

4. Faculty and Postdoc Needs and Contributions

Team Composition:

The project team is interdisciplinary in nature, bringing together expertise from three colleges: the Roy G. Perry College of Engineering, the College of Arts and Sciences, and the College of Juvenile Justice. This diverse collaboration enables a comprehensive approach to addressing challenges in trustworthy AI and cyber-physical systems. As the project evolves, the team remains open and welcoming to new members from additional disciplines who can contribute to and strengthen the initiative.

College of Engineering:

- Dr. Lijun Qian (Lead): Machine Learning, AI system design, big data, wireless networks
- Dr. Dong (Co-Lead): Deep Learning, Natural Language Processing
- Dr. Pamela Obiomon (Co-Lead): Engineering leadership and integration, Bias mitigation in AI
- Dr. Xiangfang Li: AI ethics, fairness in machine learning, edge computing
- Dr. Daniel Doe: Virtual reality/Augmented reality (VR/AR), blockchain
- Dr. Lin Gong: CPS optimization, smart grid
- Dr. Annamalai Annamalai: Digital signal processing, wireless communications
- Dr. Akshay Kulkarni: Security and privacy
- Dr. Mohamed Chouikha: AI authentication
- Dr. Lin Li: Machine learning, computer vision
- Dr. Ahmed Ahmed: IoT
- Dr. Noushin Ghaffari: Bioinformatics
- Dr. Md Shuvo: Computational biology
- Dr. Lening Wang: Processing-in-memory, AI security
- Dr. Chang Duan: Robotics, control theory
- Dr. Wenhua Yang: Physics-informed machine learning
- Dr. Ramalingam Radha: Civil infrastructure

College of Arts & Sciences:

- Dr. Anne Lippert (Lead): Cognitive psychology, knowledge representation, text analysis
- Dr. Aashir Nasim: AI for student engagement and success
- Dr. Mark Tschaepe: Philosophy, critical thinking

College of Juvenile Justice:

- Dr. Camille Gibson (Lead): Societal well-being of AI, criminal justice
- Dr. Serita Whiting: Fairness in criminal justice

Expertise Matrix:

Name	Expertise
Lijun Qian	Machine Learning, AI system design, big data analysis, wireless networks
Xishuang Dong	Deep Learning, Natural Language Processing

Pamela Obiomon	Engineering leadership and integration, Bias mitigation in AI
Xiangfang Li	AI ethics, fairness in machine learning, edge computing
Daniel Doe	Virtual reality/Augmented reality (VR/AR), blockchain
Lin Gong	CPS optimization, smart grid
Annamalai Annamalai	Digital signal processing, wireless communications
Akshay Kulkarni	Security and privacy
Mohamed Chouikha	AI authentication
Lin Li	Machine learning, computer vision
Ahmed Ahmed	IoT
Noushin Ghaffari	Bioinformatics
Md Shuvo	Computational biology
Lening Wang	Processing-in-memory, AI security
Chang Duan	Robotics, control theory
Wenhua Yang	Physics-informed machine learning
Ramalingam Radha	Civil infrastructure
Anne Lippert	Cognitive psychology, knowledge representation, text analysis
Aashir Nasim	AI for student engagement and success
Mark Tschaepe	Philosophy, critical thinking
Camille Gibson	Societal well-being of AI, criminal justice
Serita Whiting	Fairness in criminal justice

Capacity Gaps & Hiring Plans:

Postdoc Positions Needed:

Two postdocs will be recruited. It is expected that both postdocs engage in a balanced portfolio of research and teaching activities that foster collaboration across disciplines. Both postdocs will engage in research (75%) and teach one interdisciplinary course per year (25%).

1. Postdoc in AI and CPS: Focused on explainable AI and applications of trustworthy AI in cyber-physical systems;
2. Postdoc in AI Ethics: Focused on fairness auditing, transparency, and policy compliance of AI.

5. Support and Resources

Existing Infrastructure:

- CREDIT Research Center: $8M in new funding for generative AI research, NVIDIA A100 computing clusters and workstations, IBM AC922 inference servers, Dell data storage servers
- NSF supported Smart Grid Testbed: Real-time solar/wind power grid and edge computing testbeds
- Dedicated AI/ML/Big Data Lab (Room 364/365) and office space in ENCARB building

Institutional Commitments:

- Research support from the Office of Research and Innovation
- Cross-college access to AI datasets, compliance training, and data management support
- Administrative and IT support from the College of Engineering

Cross-Unit Coordination:

Bi-monthly steering committee meetings (Qian, Obiomon, Lippert, Gibson, postdocs)
Annual evaluation retreats and semi-annual reporting across units

6. Budget Outline

Budget Request:

- 2 Postdoctoral Fellows @ $60,000/year: $120,000
- Research Supplies, Computing Access, Publication Fee, Conference Travel: $10,000
- Total: $130,000

Cost-Sharing:

- College of Engineering: $10,000 in-kind (lab access, admin support)
- Matching: Additional proposals pending with NSF for scalable pilot expansion

7. Evaluation and Metrics

Success Indicators:

- 10 peer-reviewed publications over two years
- $2M in new external research funding applications
- 2 interdisciplinary graduate courses launched
- 10 students mentored (graduate and undergraduate)

Timeline & Milestones:

- Fall 2025: Postdocs recruitment and onboarding; project launch
- Spring 2026: Curriculum development and pilot evaluations
- Summer 2026: Midpoint review
- Fall 2026: Research proposal submissions and initial publications

Continuous Improvement:

Quarterly reporting, feedback from stake holders, community partners and advisory committee, and project steering oversight.