STATEMENT OF PURPOSE **Youzhe Song**
Homepage
Computer Science PhD Applicant, Fall 2026

I. INTRODUCTION: MY RESEARCH INTERESTS AND PERSPECTIVE

Current AI models have achieved remarkable success across tasks, yet their internal mechanisms remain largely a "black box." We know what models can do, and to some extent how they do it in specific details, but we lack an intuitively human-understandable view of their cognitive and reasoning processes. I believe that a key step toward more general and reliable AI lies in understanding and designing the model's internal information-processing mechanisms. My research interests center on two core questions: First, can we construct within the model a **human-intuitive, stepwise, and organic reasoning trajectory**? Second, how can we enable the model to **understand and align information from different sources** (e.g., images and text) in a human-like manner?

My goal is to explore how to realize the fusion of these two abilities in a high-dimensional latent space, enabling models not only to "see" and "read" the world, but also to reason about it in a transparent and trustworthy way.

II. ROBUST FEATURE REPRESENTATIONS (COREFACE)

My research journey began with a deep fascination for self-supervised contrastive learning. It does not rely on expensive human annotations; instead, it learns from the structure inherent in data. Its elegance and potential captivated me. However, the realities of a lab with very limited compute and diverse student directions pushed me to give it up and pursue a more fundamental alternative.

I found face recognition to be an ideal "sandbox." I observed its deep kinship with contrastive learning: both carefully shape the geometry of feature distributions in high-dimensional space and share the same goal the to be generalized in open-set. Building on this, I proposed the CoReFace framework, which is an instance of explicit design of internal mechanisms. In this fine-grained recognition task, traditional image augmentations can damage identity information. A key insight I had was that Dropout is essentially a **feature-level stochastic perturbation** that can provide the necessary views for contrastive learning without corrupting image semantics. After introducing contrastive learning, I further enhanced the weak supervising signal to make it work and rearange the pair construction in contrastive learning to avoid semantical repeat signal problem in the unified training. All of these made the framework actively **regularized** the geometric structure of the feature space. Ultimately, this approach **increased the similarity margin between positive and negative pairs by 15%**. Throughout, I led the full lifecyclefrom problem identification and design to independent validation and writingshifting from "student" to a truly **independent researcher** [1].

III. A UNIFIED FRAMEWORK FOR HETEROGENEOUS DATA (QGFACE)

After CoReFace, I yearned for research with more immediate real-world relevance. This was reinforced when I scrolled through my phone's photo album: the built-in face clustering struggled with family photos. I realized that low-quality data arising from composition, lighting, and other real-world factors differ markedly from our common datasets.

QGFace's **single-encoder architecture** is my solution. Its an internal **information dispatch system**. It simulates a "routing" process: high-quality data follow a classification path, low-quality data follow a contrastive path, and gradient truncation prevents contamination across paths. To learn to recognize low-quality data, I applied augmentation on images this time. The first thing is to design the augmentation pipeline to avoid model collapse during training. Later when contrastive learning underperformed due to insufficient positive pairs, I designed a **Proxy-Updated Real-Time Queue** that boosted performance on low-quality data. Racing against a deadline, I structured my days into intense cycles; GPU time became my rest time. Far from being a burden, I felt little depression at that time as I knew that this was something being built. It proved that what drives me is not external expectation, but the intrinsic intellectual joy of solving hard problems. Ultimately, QGFace reached SOTA on low-quality datasets with **only a 0.3% trade-off** on high-quality data, validating that my research philosophy can yield elegant, robust, and practical systems [2].

IV. EXPLORATION: CLARIFYING MY DIRECTION THROUGH WORKING PRACTICE

Despite some academic progress, repeated setbacks and uncertainty in the submission process led to a period of deep confusion and self-doubt. To find clarity, I decided to step into industry and **stress-test** my intrinsic motivation.

I first joined KeyoneAI, a startup led by former IBM China Chief AI Architect Jie Fang. There, I translated frontier generative AI into product, felt the pulse of rapid iteration, advocated technology, and engaged with users.

Later, I served as the sole technical lead on the organizing committee of the Worldwide Educators Conference (WWEC) [3]. Beyond ensuring system reliability, I accelerated the team's digital transformation, developed internal tools to generate 1000+ complex posters, and connected 3,000 attendees, 1,000 exhibitors, 10+ vendors, and many ad-hoc sessionsworking 80 hours per week for nearly three months. While challenging and rewarding, these experiences still could not replace the pure intellectual excitement and flow I feel in research, witnessing breakthroughs in a field and contributing to them. This deliberate detour granted me unprecedented clarity: my deepest drive is to question fundamentals, refute and rebuild existing solutions, and innovate effectively across disciplines.

More importantly, this exploration helped me reframe what used to be excessive self-scrutiny into a unique research lens. I understand deeply that **a truly intelligent system is not defined by flawless unidirectional reasoning but by its capacity to handle internal conflict, self-examination, and iterative revision**the essence of human reflection and "productive struggle."

V. FUTURE: ALIGN SEMANTICS & REASON IN LATENT SPACE

I now plan my Ph.D. research with renewed clarity and conviction. I aim to combine my experience in **mechanism design** with reflections on **human cognition**, focusing on two capabilities of large models:

1. Building a Unified Semantic Space: My first goal is to study how to effectively map diverse information into a unified, interpretable latent space. This goes beyond cross-modal mapping. It is the lower code of how model understand the world laws, like physics. I believe that efficient cross-modal semantic alignment is a first step toward foundational models that can comprehensively understand the world.

2. Realizing Structured Reasoning in Latent Space: The other goal is to design *structured*, chain-of-thought-like reasoning trajectories within the latent space. I want models not only to be like the early computers that used punched paper tape for data input and output (which is transformer doing), but also to present a decomposable and traceable reasoning process in itself.

In summary, my research proceeds on two fronts: through **sementic alignment**, enabling models to "see" a richer world; and through **latent reasoning**, enabling them to "think" more clearly and logically.

VI. CONCLUSION

With hands-on experience designing internal mechanisms and a clear plan to integrate *semantic alignment* with *latent reasoning*, I am prepared for the challenges of a Ph.D. I look forward to contributing to the next generation of more capable and trustworthy AI systems in a creative and supportive environment.

REFERENCES

[1] Youzhe Song and Feng Wang. Coreface: Sample-guided contrastive regularization for deep face recognition. *Pattern Recognition*, 152:110483, 2024.

[2] Youzhe Song and Feng Wang. Quality-guided joint training for mixed-quality face recognition. In *2024 IEEE International Conference on Automatic Face and Gesture Recognition (FG)*. IEEE, 2024.

[3] Worldwide educators conference (wwec). `https://www.wwec820.com/`.