

Kun Xu

Ph.D. Candidate in Cyberspace Security

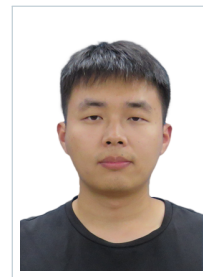
Trustworthy AIGC, Generative Model Security, and Multimedia Forensics

xukun930@nuaa.edu.cn | +86 181 1063 3068 | xukun12138.github.io

Google Scholar | ORCID | GitHub | DBLP | OpenReview

Nanjing, Jiangsu 211106, China | Born Sep. 1998 in Lu'an, Anhui, China

Last updated: August 2026



Profile

Ph.D. candidate at Nanjing University of Aeronautics and Astronautics (NCAA), advised by Prof. Yushu Zhang, working on the safety, evaluation, and reliability of generative AI. Research spans concept-level risks in text-to-image systems, malicious and toxic content detection, uncertainty and calibration in diffusion models, privacy-aware synthesis, and deepfake forensics. Author of five published or accepted papers, including two papers at CCF-A venues, with six additional manuscripts under review or submission. Recipient of a China Scholarship Council award supporting a one-year research visit in Italy.

Research Interests

- **Trustworthy AIGC and foundation-model evaluation:** concept-level safety, risk assessment, calibration, toxic and malicious content detection, and reliable governance of generative systems.
- **Diffusion-model uncertainty and representation:** uncertainty quantification, representation consistency, latent-space constraints, and generation-path guidance for personalized text-to-image models.
- **Privacy and multimedia forensics:** privacy-aware portrait release, deepfake detection and localization, adversarial defense, source-camera forensics, and medical-image integrity.

International Research Experience

Sep. 2025 – Aug. 2026	Visiting Ph.D. Researcher, University of Insubria	Italy
Funded by the China Scholarship Council Excellence Talent Programme (International Training Programme for Outstanding Engineers). Supervised by Prof. Elena Ferrari (IEEE/ACM Fellow), with research collaboration involving Prof. Pierangela Samarati (IEEE/ACM Fellow) and Prof. Vincenzo Piuri (IEEE Fellow), University of Milan.		

Education

Sep. 2023 – Oct. 2026 (expected)	Ph.D. in Cyberspace Security	Nanjing, China
College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics. Advisor: Prof. Yushu Zhang.		
Sep. 2020 – Jun. 2023	Master's in Computer Science and Technology	Huainan, China
School of Computer Science and Engineering, Anhui University of Science and Technology. Advisor: Prof. Gaoming Yang.		
Sep. 2016 – Jun. 2020	Bachelor's in Computer Science and Technology	Huainan, China
School of Computer Science and Engineering, Anhui University of Science and Technology.		

Publications and Accepted Papers

1. **Kun Xu**, Wenying Wen*, Shuren Qi, Tao Wang, Yushu Zhang, and Yuming Fang. "Detecting Malicious Concepts Without Image Generation in AI-Generated Content (AIGC)." *IEEE Transactions on Dependable and Secure Computing*, 23(4):8699–8716, 2026. CCF A
2. **Kun Xu**, Yushu Zhang*, Tao Wang, Shuren Qi, Barbara Carminati, Elena Ferrari, and Yuming Fang. "Concept-Level Risk and Calibration for Governance in Diffusion Foundation Models." In the *34th ACM International Conference on Multimedia (ACM MM)*, 2026. Accepted. CCF A

3. **Kun Xu**, Shuren Qi, Yushu Zhang*, Wenying Wen, and Hua Zhang. "A Facial Manipulation Adversarial Defense Approach for Image Post-Processing." *Computer Engineering & Science*, 47(12):2181–2194, 2025. In Chinese. CCF T2
4. **Kun Xu**, Gaoming Yang*, Xianjin Fang, and Ji Zhang. "Facial Depth Forgery Detection Based on Image Gradient." *Multimedia Tools and Applications*, 82(19):29501–29525, 2023. CCF C
5. Gaoming Yang, **Kun Xu***, Xianjin Fang, and Ji Zhang. "Video Face Forgery Detection via Facial Motion-Assisted Capturing Dense Optical Flow Truncation." *The Visual Computer*, 39(11):5589–5608, 2023. CCF C

* Corresponding author, as marked in the source publication list.

Manuscripts Under Review or Submission

1. **Kun Xu**, Yushu Zhang*, Tao Wang, Shuren Qi, Zhongyun Hua, and Yuming Fang. "Understanding Concept-Driven Diffusion Model with Uncertainty." Manuscript under review at the *International Journal of Computer Vision*. CCF A
2. **Kun Xu**, Yushu Zhang*, Shuren Qi, Tao Wang, Barbara Carminati, Elena Ferrari, and Yuming Fang. "Identity Replacement Is Not Anonymization: Coupled-Condition Diffusion for Privacy-Aware Portrait Release." Manuscript under review at the *Network and Distributed System Security Symposium (NDSS)*. CCF A
3. **Kun Xu**, Shuren Qi, Qing Tan, and Yushu Zhang*. "Source-Camera-Aware Forgery Localization for IoT Visual Sensing from Photo-Response Non-Uniformity." Manuscript under review at the *Journal of Computer Science and Technology*. CCF B
4. Wenying Wen, Zhen Chen, **Kun Xu**, Xiaojie Tao, and Yushu Zhang. "Detecting Subtle Lesion Tampering in Medical Images: A Deep Hashing Method with Multi-Scale Sparse Attention and Key-Weighted Mechanism." Manuscript under review at *IEEE Transactions on Information Forensics and Security*. CCF A
5. Ziyue Yuan, Wenying Wen, **Kun Xu**, Tao Wang, and Yushu Zhang. "Having the Cake and Eating It Too: Differential-Response Hashing for Unified Integrity Authentication of Medical Images and Metadata." Manuscript under review at *IEEE Transactions on Image Processing*. CCF A
6. Tao Wang, Yushu Zhang, Xiangli Xiao, **Kun Xu**, Lin Yuan, Wenying Wen, and Yuming Fang. "Rethinking Privacy Protection for Subject Faces in Photo Sharing: From Perturbation-Based to Synthesis-Based Methods." Manuscript under review at *IEEE Transactions on Information Forensics and Security*. CCF A

Research Projects

2021 – 2024	National Key R&D Programme of China	Participant
	"Blockchain-Governed Blockchain": cross-chain consensus algorithms and consistency proofs.	
2023 – 2026	Jiangxi Provincial Outstanding Young Scientists Fund Project	Participant
	Key technologies for data security in multimedia Internet of Things systems.	
2021 – 2023	High-End Foreign Experts Project, Ministry of Science and Technology	Participant
	Machine-learning-based adversarial attacks and defenses.	
2022 – 2023	Innovation Project, Anhui University of Science and Technology	Principal Investigator
	Detection and identification of deepfake face images and videos.	

Honors and Awards

2025	Excellence Talent Programme National Scholarship	
	China Scholarship Council (CSC)	
2024	National-Level Award, 3rd National College Student Data Analysis Popular Science Competition	
2023 – 2025	Graduate Academic Scholarship	
	Nanjing University of Aeronautics and Astronautics	
2021 – 2023	Graduate Academic Scholarship	
	Anhui University of Science and Technology	

Honors and Awards (continued)

2023	Outstanding Graduate of Anhui Province
	Anhui Provincial Department of Education
2023	Outstanding Graduate
	Anhui University of Science and Technology

Academic Service and Memberships

Program committees	ADFM 2026 @ CVPR; PEARC '26; SiMLA '26.
Conference reviewing	ACM Multimedia 2026; ADFM 2026.
Journal reviewing	Computer Vision and Image Understanding; Signal Processing; Journal of Computer Science and Technology; Multimedia Systems; Expert Systems; and Scientific Reports.
Memberships	ACM Emerging Interest Group on Trustworthy and Responsible Systems (EIGTRUST); China Society of Image and Graphics (CSIG).

Selected Talks and Research Exchanges

Nov. 2025	Research exchange at the University of Milan <i>Milan, Italy</i>
	Academic discussion with Prof. Pierangela Samarati and Prof. Vincenzo Piuri on trustworthy AI and generative-model security.
Oct. 2025	Minisymposium: Generative Concept Security in Trustworthy AIGC <i>University of Insubria, Italy</i>
	Presentation connecting concept-level safety, uncertainty, and trustworthy generative systems.

Technical and Research Skills

Research	Trustworthy AI, foundation-model evaluation, generative-model security, uncertainty quantification, deep learning, computer vision, multimedia forensics, and cyberspace security.
Programming	Python and PyTorch; working knowledge of Java and C/C++.
ML and data tools	NumPy, pandas, scikit-learn, Matplotlib, and OpenCV.
Systems and workflow	Linux, Git, Conda/Anaconda, reproducible experimentation, and scholarly writing.
Mathematical methods	Probability, linear algebra, discrete mathematics, numerical analysis, and derivation of methods from the research literature.
Languages	English and Mandarin Chinese.

Publication and manuscript status is current as of August 2026. CCF labels follow the classifications supplied in the source CV.