CV

Research experience, education, selected outputs and technical expertise.

Basics

Name	Yu Zhang
Label	PhD Researcher and Research Fellow
Email	yu.zhang@surrey.ac.uk
Url	https://yurbro.github.io

Summary PhD researcher in Chemical and Process Engineering at the University of Surrey, working at the interface of AI, digital twins and data-driven engineering. Research spans data-efficient formulation and drug-delivery design, uncertainty-aware modelling, scientific evidence extraction and reproducible machine learning with real-world clinical data. Also contributes to battery intelligence research through data processing, computational algorithms and research software.

Work

2026.09 - PRESENT

Research Fellow — Advanced Battery Intelligence and Technology Development

Advanced Technology Institute (ATI), Guildford, UK

University of Surrey

Work with Dr Kai Yang on an EPSRC IAA project in battery intelligence, contributing data processing, algorithm development and research software workflows.

- Process and organise experimental datasets for computational analysis.
- Contribute to algorithm development and data-driven research workflows.
- Develop research software to support analysis and workflow automation.

2026.06 - PRESENT

Research Assistant — Clinical Machine Learning

Guildford, UK

University of Surrey

Work with Dr Xilu Wang (University of Surrey) and Dr Mingkun Tan (University of Münster, Germany) on reproducible machine-learning workflows for real-world clinical data in assisted reproduction and IVF.

- Define clinically meaningful outcomes and construct analysis-ready features from heterogeneous, longitudinal treatment records.
- Address incomplete records and outcome availability when preparing data for patient-level outcome modelling.
- Design patient-aware and time-aware validation to reduce information leakage and assess model generalisation.
- Maintain auditable workflows with attention to confidentiality, data provenance and reproducibility.

2024 - PRESENT

Teaching Assistant

Guildford, UK

University of Surrey

Teaching support across engineering systems, science principles and advanced pharmaceutical formulations.

- ENG2128 Engineering Systems and Management (2024–25) · Lecturer: Dr Lei Xing
- ENG1080 Science Principles for Engineers (2025–26) · Lecturer: Prof. Madeleine Bussemaker
- CHEM042 Advanced Pharmaceutical Formulations (2025–26 and 2026–27) — Teaching support in formulation science, pharmaceutical product design and advanced delivery systems. · Lecturer: Prof. Tao Chen

2023.09 - PRESENT

PhD Researcher — Machine Learning and Digital-Twin Modelling

Guildford, UK

University of Surrey

Data-efficient modelling and experimental decision-making for pharmaceutical formulation and drug delivery.

- Develop Gaussian-process surrogate models, Bayesian optimisation and active-learning workflows for adaptive and multi-objective experimental design.
- Model drug-release and permeation behaviour using domain-knowledge-constrained symbolic regression and temporal prediction from partial experimental trajectories.
- Develop reproducible LLM-enabled workflows for scientific evidence extraction and dataset construction.

2021.05 - 2021.06

Technical Lead — Big Data AI Analysis System for Shale Gas Trading Market

China

External client project

Python software for energy-market data collection, analysis and visualisation.

- Built automated collection and analysis of real-time and historical market data.
- Delivered executable software to an external client.

2020.12 - 2022.07

Research Assistant — Data-Driven Reservoir Simulation Analysis

Beijing, China

China University of Petroleum - Beijing

Data-driven analysis of numerical reservoir-simulation outputs.

- Used PCA/MPCA, Hotelling's T^2, SPE and self-organising maps to extract spatial and temporal features.
- Classified dominant flow regimes from simulation outputs.

2020.12 - 2021.10

Research Assistant — Intelligent Monitoring of Electric Submersible Pump Systems

Beijing, China

National Key R&D Programme of China project

Data-driven condition monitoring, fault classification and early warning from multivariate operational and process data.

- Applied PCA/SFA, regression, clustering and SVM methods to operational sensor and process data.
- Implemented Python-based analytical software for condition monitoring and fault diagnosis.

Education

2026.06 - 2026.08

Kraków, Poland

Visiting PhD Researcher

Jagiellonian University Medical College

Faculty of Pharmacy, Department of Social Pharmacy · Non-degree research placement under COST Action CA21108 (NETSKINMODELS)

Host: Prof. Barbara Wiśniowska · Collaborator: Dr Laura Krumpholz

- Research visit: 15 June – 14 August 2026; home PhD institution: University of Surrey.
- SkinPhysio-Extract: LLM-assisted extraction, provenance auditing and verification of quantitative skin physiology evidence.

2023.09 - PRESENT

Guildford, UK

PhD

University of Surrey

Chemical and Process Engineering

Supervisors: Prof. Tao Chen , Assoc. Prof. Dimitrios Tsaoulidis

- Expected thesis submission: late 2026 / early 2027.
- Digital twin and AI methods for topical formulation design

2020.09 - 2023.07

Beijing, China

MEng

China University of Petroleum - Beijing

Control Science and Engineering, College of Artificial Intelligence

Supervisor: Prof. Xiaoyong Gao

Research focus: Data-driven fault monitoring and diagnosis of electric submersible pumps under nonlinear dynamics and limited fault samples.

- Developed LPP-enhanced dynamic kernel PCA for nonlinear process monitoring and fault identification.
- Developed improved conditional GAN-based data augmentation for fault diagnosis using extremely limited samples.
- Investigated optimisation-based semi-supervised learning for multi-class fault diagnosis using unlabeled industrial data.

Related publications: The Canadian Journal of Chemical Engineering (2023) · Journal of the Franklin Institute (2024)

2015.09 - 2019.07

Sichuan, China

BEng

Panzhihua University

Automation, School of Intelligent Manufacturing

Publications

2026

A modelling-based comparison of IVRT and synthetic-membrane permeation for early formulation screening

RSC Pharmaceutics

Quantifies alignment between IVRT and synthetic-membrane permeation assays.

2026

Confidence-Bound Early Stopping of Experiments with Sequential Calibration

Chemical Engineering Research and Design, 230, 756-766

Early stopping with calibrated control of the false-stop rate.

2026

Autonomous AI-Driven Design for Skin Product Formulations

Advanced Intelligent Discovery, e70100

A roadmap for assay-aware, closed-loop and increasingly autonomous skin formulation development.

2026

Active learning-based adaptive optimisation for developing dermal drug formulations

Chemical Engineering Research and Design, 232

Adaptive batch allocation between expected-improvement and information-seeking formulation candidates.

2026

Domain Knowledge Constrained Symbolic Regression for Optimising Dermal Drug Formulations

Pharmaceutical Research

Domain-knowledge-constrained symbolic regression for interpretable modelling and optimisation of dermal drug formulations.

2026

Optimisation of Topical and Transdermal Pharmaceutical Formulations: Methods, Challenges and Opportunities

SmartSys, 2(4), e70047

Yongrui Xiao, Yu Zhang, Chunlin Chen, Dimitrios Tsaoulidis, Kai Yang and Tao Chen. Published online on 6 November 2026. Review of statistical design, machine learning, Bayesian optimisation and emerging autonomous laboratories for topical and transdermal formulations.

2026

SkinMiner: LLM-powered data mining to accelerate formulation design for skin products

SSRN preprint (under review at International Journal of Pharmaceutics)

Yu Zhang, Xilu Wang, Dimitrios Tsaoulidis and Tao Chen. LLM-powered extraction, verification and structuring of dermal formulation evidence. Preprint; peer review ongoing.

2026

A Spectral-Continuity-Guided Bayesian Learning Framework for Infrared Spectroscopy-Based Soft Sensor Modeling

Measurement

Pengcheng Wu, Tao Chen, Yu Zhang, Xiaobo Zou and Haoran Li. Collaborative manuscript on spectral-continuity-guided Bayesian learning for infrared spectroscopy-based soft sensor modelling. Published online in Measurement in 2026.

2025

An early decision-making algorithm for accelerating topical drug formulation optimisation

Computers & Chemical Engineering, 201, 109224

Probabilistic early decisions from partial IVPT trajectories.

2024

Data augmentation using improved conditional GAN under extremely limited fault samples and its application in fault diagnosis of electric submersible pump

Journal of the Franklin Institute, 361(4), 106629

Virtual sample generation for fault diagnosis under extreme data scarcity.

2023

Improved dynamic kernel PCA based on local preserving projections and its application for electric submersible pump fault diagnosis

The Canadian Journal of Chemical Engineering, 101(8), 4539-4554

Local-structure-preserving nonlinear dynamic process monitoring.

2022

Green scheduling optimization of shuttle tanker route considering carbon emissions

Chemical Industry and Engineering, 39(2), 23-31

Multi-objective route scheduling with carbon-emission considerations.

Presentations

4 AUG 2026

LLM-powered multimodal agentic data mining for dermal formulation science

Zhang, Y.; Tsaoulidis, D.; Chen, T.

28th International Congress of Chemical and Process Engineering (CHISA 2026) · Prague, Czech Republic

Oral presentation.

Projects

2026.09 - PRESENT

Battery Intelligence

Data processing, computational algorithms and research software for an EPSRC IAA battery intelligence project at Surrey's Advanced Technology Institute, working with Dr Kai Yang.

- Experimental data processing
- Algorithm development
- Research software

2024 - 2026

Adaptive Batch Optimisation

Active learning for formulation optimisation under a limited per-batch experimental budget.

- Gaussian process regression
- Expected improvement
- Hypervolume contribution

2023 - 2025

Early Decision-Making Algorithm

Forecasting final permeation outcomes from partial IVPT trajectories.

- Uncertainty quantification
- Probability of exceedance
- Experimental resource saving

25 - PRESENT

Domain-Constrained Symbolic Regression

Interpretable dynamic modelling and optimisation of dermal drug release.

- Domain constraints
- Dynamic modelling
- Interpretable optimisation

2026 - PRESENT

SkinMiner

LLM-powered extraction and verification of dermal formulation evidence. SSRN preprint available; under review at International Journal of Pharmaceutics.

- Scientific LLMs
- Evidence provenance
- Structured datasets

2026.06 - PRESENT

Clinical Machine Learning for Assisted Reproduction

Ongoing research on reproducible outcome-modelling workflows for real-world IVF data.

- Outcome definition and feature construction
- Patient-aware and time-aware validation
- Data provenance and reproducibility

Awards & Grants

2026.06

Short-Term Scientific Mission and Mobility Funding

COST STSM, UK Turing Scheme and FEPS

€2.8K COST STSM, £1K UK Turing Scheme and £1K FEPS funding for a June–August 2026 research placement at Jagiellonian University, Poland. Hosted by Prof. Barbara Wiśniowska and Dr Laura Krumpholz

2025.01

UK Turing Scheme Research Placement

UK Turing Scheme

£1K mobility funding for a January-March 2025 research placement at West China Hospital, Sichuan University.

2023.09

China Scholarship Council - University of Surrey Award

China Scholarship Council and University of Surrey

Full-time PhD scholarship, September 2023-September 2027. Grant No. 202306440060.

2022

Outstanding Graduate Student (优秀研究生)

Beijing Automation Society (北京自动化学会)

Awarded the 2022 Outstanding Graduate Student distinction.

Skills

Machine Learning and Modelling

Gaussian processes
Random forests
SVR / SVM
Neural networks
PCA / MPCA
Slow feature analysis
Self-organising maps

Optimisation and Decision-Making

Bayesian optimisation
Active learning
Uncertainty quantification
Temporal prediction
NSGA-II
Symbolic regression

Clinical Data and Validation

Data-quality auditing
Outcome definition
Feature construction
Longitudinal data
Patient-aware and time-aware validation

Formulation and Experiments

Dermal drug delivery
IVRT
IVPT
Formulation design
Digital twins
Automated experimentation

LLMs and Scientific Evidence

Multimodal and agentic data mining
Evidence extraction
Scientific workflow automation
Evidence provenance

Scientific Computing

Python
PyTorch
TensorFlow
MATLAB
pandas
scikit-learn
SciPy
Matplotlib
Git / GitHub
Reproducible research

Languages

Chinese

Native speaker

English

Professional working proficiency

Interests

Research interests

AI for engineering and science
Digital twins
Formulation and drug delivery
Clinical machine learning
Scientific data mining
Process monitoring and fault diagnosis

Certificates

Microsoft Certified: Azure AI Fundamentals

Microsoft 2021

Memberships

Institution of Chemical Engineers (IChemE) · Affiliate member

Chinese Association of Automation (CAC) · Member

Chemical Industry and Engineering Society of China (CIESC) · Member

Academic Service

Journal peer review

- **Journal of Automation and Intelligence** — 6 completed reviews recognised for October 2025 – September 2026.
- **Franklin Open** — 7 completed reviews recognised for October 2025 – April 2026.

Conference reviewing

- **China Automation Congress (CAC)** — Reviewer recognition in 2023, 2024 and 2025.

Conference activities and photo journal