

# 从人文材料到智能洞见 · 「吾与点」数字人文工作坊

## From Humanities Materials to AI Insights · "Wu Yu Dian" Digital Humanities Workshop

### 一、工作坊基本信息 I. Overview

本工作坊以“从人文材料到智能洞见”为主题，由北京大学数字人文研究中心主办，面向人文学科研究者，系统介绍「吾与点」智能数据平台与「识典古籍」古籍智能整理平台的核心功能与研究应用，帮助学员在 2.5 天的密集课程中，完整体验从原始人文材料到智能研究助手的全流程实践，赋能人工智能时代的人文研究新范式。

「吾与点」是北京大学建数字人文研究中心开发的智能数据平台，旨在为人文学科研究者提供从原始材料到智能体的一站式解决方案。「识典古籍」是北京大学与字节跳动联合开发的古籍智能整理平台，提供从 OCR 识别、自动句读、专名识别到自动校勘的全流程的古籍智能整理数字化服务。本工作坊为期两天半，课程涵盖古籍智能整理、知识图谱理论与构建、多模态智能数据抽取、知识库与智能体搭建，以及可视化网页智能开发。课程通过理论讲解与动手实操相结合的方式，帮助学员掌握以 AI 赋能人文研究的完整 workflow，让研究者无需编程即可完成从数据提取到成果呈现的完整闭环。

Themed "**From Humanities Materials to AI Insights**", this workshop is hosted by the **Research Center for Digital Humanities at Peking University** and **designed for researchers in the humanities**. It offers a systematic introduction to the core features and research applications of two platforms: **Wu Yu Dian** (吾与点), an intelligent data platform, and **Shidian Guji** (识典古籍), an AI-powered platform for classical text editing. Across 2.5 days of intensive instruction, participants will work through **the full pipeline from raw humanities materials to AI-powered research assistants**, gaining hands-on experience with **a new paradigm for humanities research in the age of AI**.

**Wu Yu Dian**, developed by the Research Center for Digital Humanities at Peking University, is an intelligent data platform that provides humanities researchers with a one-stop solution—from raw materials all the way to deployable AI agents. **Shidian Guji**, jointly developed by Peking University and ByteDance, is an AI-powered platform supporting the full classical text editing workflow, including OCR, automatic punctuation, named entity recognition, and automatic textual collation. The 2.5-day curriculum covers **AI-assisted classical text editing, knowledge graph construction, multimodal data extraction, knowledge database creation, AI agent building, and AI-assisted development of interactive web visualizations**. Combining lectures with hands-on practice, the workshop equips participants with a complete AI-empowered workflow for humanities research—closing the loop from data extraction to final presentation, all without writing a single line of code.

- **时间 / Date:** 2026 年 7 月 22 日 (下午) – 7 月 24 日 / July 22 (afternoon) – July 24, 2026
- **地点 / Venue:** 韩国梨花女子大学 / Ewha Womans University, South Korea

- **每日时长 / Daily Hours:** 上午 9:30 – 12:00; 下午 13:30 – 17:00 / Morning 9:30 – 12:00; Afternoon 13:30 – 17:00
- **使用平台 / Platforms:**
  - 吾与点/Wu Yu Dian: <https://www.wuyudian.net/>
  - 识典古籍/Shidian Guji: <https://www.shidianguji.com/>

## 二、授课团队 II. Instructors

**王军**，北京大学信息管理系、人工智能研究院、历史学系兼职教授，北京大学数字人研究中心主任。主要从事数字人文、文化计算、古籍智能等跨领域研究，主持研发"吾与点"智能数据平台与《宋元学案》知识图谱等。

**Prof. Dr. WANG Jun** is Professor at the Department of Information Management and the Institute for Artificial Intelligence, with a joint appointment at the Department of History, all at Peking University. He serves as Director of the Research Center for Digital Humanities at Peking University. His research spans digital humanities, cultural computing, and AI for classical texts. He leads the development of the Wu Yu Dian intelligent data platform and the Song-Yuan Xue'an Knowledge Graph.

**杨浩**，北京大学人工智能研究院副研究员，北京大学数字人文研究中心常务副主任。中国哲学博士（北京大学），曾任北京大学哲学系助理教授与《儒藏》编纂研究中心研究员。学术专长涵盖宋明理学、隋唐佛学等领域，致力于推进人文学术的数字化转型，负责北大与字节公益合作研发的"识典古籍"平台的设计指导。

**Dr. YANG Hao** is Associate Research Fellow at the Institute for Artificial Intelligence, Peking University, and Deputy Director of the Research Center for Digital Humanities at Peking University. He holds a PhD in Chinese Philosophy from Peking University and previously served as Assistant Professor at PKU's Department of Philosophy and Research Fellow at the Confucian Canon Compilation and Research Center. His scholarly expertise spans Song-Ming Neo-Confucianism and Sui-Tang Buddhism, with a sustained commitment to advancing the digital transformation of humanities scholarship. He oversees the design of the Shidian Guji platform, jointly developed by Peking University and ByteDance Public Welfare.

**位通**，北京大学信息管理系助理教授，北京大学数字人文研究中心研究员，法国格勒诺布尔-阿尔卑斯大学计算机科学博士。主要从事数字人文、古籍数字化、知识图谱与文化遗产等交叉领域研究，主持国家自然科学基金青年基金项目与欧盟国际合作项目，参与多项国家古籍数字化重点工程。

**Dr. WEI Tong** is Assistant Professor at the Department of Information Management, Peking University, and Research Fellow at the Research Center for Digital Humanities at Peking University. He holds a PhD in Computer Science from Université Grenoble Alpes, France. His research sits at the intersection of digital humanities, classical text digitization, knowledge graphs, and cultural heritage.

He is the principal investigator of a Young Scientists Fund project of the National Natural Science Foundation of China and an EU international collaboration project, and contributes to several national flagship classical text digitization projects.

**潘娅婷**，北京大学信息管理系和北京大学数字人文研究中心博士研究生。研究方向涵盖数字人文、多元文化分析、多模态自然语言处理与大语言模型。苏黎世大学计算语言学与语言技术硕士，曾任苏黎世大学数字社会倡议中心研究助理。

**PAN Yating** is a PhD candidate at the Department of Information Management and the Research Center for Digital Humanities at Peking University. Her research interests include digital humanities, cross-cultural analysis, multimodal natural language processing, and large language models. She holds an MA in Computational Linguistics and Language Technology from the University of Zurich, where she previously served as Research Assistant at the Digital Society Initiative.

**王兆基**，北京大学信息管理系和北京大学数字人文研究中心博士研究生。研究方向涵盖数字人文、古汉语信息处理、大型语言模型等领域。北京师范大学文学硕士。参与多项数字人文基础设施项目，包括"吾与点智能数据平台"及"经籍指掌"典籍目录系统等的研发与维护工作。

**WANG Zhaoji** is a PhD candidate at the Department of Information Management and the Research Center for Digital Humanities at Peking University. His research covers digital humanities, classical Chinese information processing, and large language models. He holds an MA from Beijing Normal University. He contributes to several digital humanities infrastructure projects, including the development and maintenance of the Wu Yu Dian intelligent data platform and the Jingji Zhizhang classical text catalogue system.

### 三、学习目标 III. Learning Objectives

完成本工作坊后，学员将系统理解 AI 技术如何赋能人文学科研究，掌握从原始材料到智能研究助手的完整方法论与实践路径，包括使用识典古籍完成古籍 OCR 与自动句读、使用吾与点构建多语种知识图谱、从文本/PDF/图像中自动生成结构化智能表格、搭建基于个人数据的智能体研究助手，并独立制作 GIS 地图、统计图表等可视化网页应用。所有操作无需编程，重在建立"数据驱动的人文研究"思维方式。

Upon completing this workshop, participants will **systematically understand how AI empowers humanities research, and master the full methodology and practical path from raw materials to intelligent research assistants**. This includes: using **Shidian Guji** to perform OCR and automatic punctuation on classical texts; using **Wu Yu Dian** to construct multilingual knowledge graphs; automatically generating structured intelligent datasheets from text, PDFs, and images; building AI agent research assistants on top of personal data; and independently creating web applications such as GIS maps and statistical charts. **All operations require no coding. The focus is on cultivating a "data-driven humanities research" mindset.**

#### 3.1 你将学到的核心概念 Core Concepts You Will Learn

- **古籍数字化全流程 (Classical Text Digitization)** : 识典古籍平台的 OCR、自动句读、版

本校勘 / OCR, automatic punctuation, and version collation on the Shidian Guji platform

- **知识图谱 (Knowledge Graph)** : 知识图谱的理论基础与构建方法 / Theoretical foundations and construction methods of knowledge graphs
- **多模态大语言模型 (Multimodal LLMs)**: LLM 如何处理文本、图像、PDF 等多模态数据 / How LLMs process text, images, and PDFs as multimodal inputs
- **智能数据表格与提示词工程 (Intelligent Datasheets & Prompt Engineering)**: 将研究问题翻译成字段结构, 驱动 AI 完成结构化数据生成 / Translating research questions into field structures, then prompting AI to generate structured data
- **数据集与语义检索 (Datasets & Semantic Retrieval)**: 向量嵌入与跨语言、跨表述的语义检索原理 / Vector embeddings and cross-lingual, cross-phrasing semantic search
- **智能体构建 (AI Agent Construction)** : 基于自有知识库构建专属研究助手 / Building a personal research assistant grounded in your own knowledge base
- **数据可视化与网页应用 (Data Visualization & Web Applications)** : Vibe Coding 生成 GIS 地图、统计图表 / Vibe-coding GIS maps and charts via natural language

## 3.2 你将带走的成果 Take-Home Outcomes

工作坊结束时, 每位学员将拥有一套基于**个人材料**的完整研究资产: 一份经识典古籍数字化处理的文本、一张智能数据表、一个可检索的数据集、一个能问答的智能体, 以及一个可视化网页应用 Demo。

By the end of the workshop, each participant will own a complete suite of research assets built from their **own materials**: a digitized classical text processed via Shidian Guji, an intelligent datasheet, a searchable dataset, a question-answering AI agent, and a visualization web application demo.

## 四、课程安排 IV. Schedule

 7月22日 (周三): 开幕式 + Workshop 导览·识典古籍

### Wednesday, July 22: Open Ceremony + Workshop Overview · Shidian Guji

时间 Time	内容 Content
9:30 – 10:00	开幕式 (线上) Opening Ceremony (Online)
10:00 – 10:15	茶歇 Coffee break
10:15 – 10:45	培训班整体安排介绍 Workshop overview
10:45 – 11:00	茶歇 Coffee break
11:00 – 12:00	识典古籍平台介绍 Introduction to Shidian Guji Platform
12:00 – 13:30	午餐 Lunch
13:30 – 14:30	识典古籍：古籍整理全流程讲解 Shidian Guji: Workflow for Processing and Organizing Classical Texts
14:30 – 14:45	茶歇 Coffee break
14:45 – 15:45	识典古籍：古籍整理全流程讲解 Shidian Guji: Workflow for Processing and Organizing Classical Texts
15:45 – 16:00	茶歇 Coffee break
16:00 – 17:00	学员实操：学员跟练古籍数字化 Hands-on practice: digitizing classical texts on Shidian Guji

## 7月23日（周四）：知识图谱·智能表格

### Thursday, July 23: Knowledge Graphs · Intelligent Datasheets

时间 Time	内容 Content
09:30 – 10:30	知识图谱理论 Knowledge graph theory
10:30 – 10:45	茶歇 Coffee break
10:45 – 12:00	吾与点：知识图谱构建实操 Wu Yu Dian: hands-on knowledge graph construction
12:00 – 13:30	午餐 Lunch
13:30 – 14:30	吾与点智能表格：古籍文本结构化数据抽取 Wu Yu Dian intelligent datasheets: Extract structural data from classical text
14:30 – 14:45	茶歇 Coffee break
14:45 – 15:45	多模态智能表格：图片和 PDF 结构化数据抽取 Multimodal datasheets: Extract structural data from images and PDFs
15:45 – 16:00	茶歇 Coffee break
16:00 – 17:00	学员实操：基于个人材料构建智能表格 Hands-on practice: building intelligent datasheets from your own materials

## 7月24日（周五）：数据集·智能体·网页应用·结业

### Friday, July 24: Datasets · AI Agents · Web Applications · Closing

时间 Time	内容 Content
09:30 – 10:10	数据集与知识库：实现语义检索 Datasets & knowledge bases: enabling semantic search
10:10 – 10:25	茶歇 Coffee break

10:25 – 11:25	<b>智能体：问答与研究助手（核心概念和构建流程）</b> AI Agents: Q&A and research assistants (core concepts and building pipeline)
11:25 – 11:40	茶歇 Coffee break
11:40 – 12:00	<b>网页应用入门：统计分布、GIS 地图</b> Intro to web applications: statistical distributions and GIS maps
12:00 – 13:30	午餐 Lunch
13:30 – 14:30	<b>学员实操（一）：基于个人材料 — 数据集 + 智能体</b> Hands-on practice (1): your own materials — datasets + AI agents
14:30 – 14:45	茶歇 Coffee break
14:45 – 15:45	<b>学员实操（二）：基于个人材料 — 网页应用构建</b> Hands-on practice (2): your own materials — building a web application
15:45 – 16:00	茶歇 Coffee break
16:00 – 17:00	<b>个人 Demo 展示与点评 + 结业总结 + 课程反馈调研</b> Individual demo presentations + faculty feedback + closing remarks + course evaluation survey

## 五、课前准备清单 V. Pre-Workshop Checklist

### 5.1 设备与账号 Devices & Accounts

- **自带电脑与充电器 Bring your own laptop and charger**
- 提前完成以下平台的账号注册（支持中、韩等手机号与邮箱注册）：  
**Register accounts on the following platforms in advance (supports Chinese, Korean phone numbers and email):**
  - 识典古籍 Shidian Guji: <https://www.shidianguji.com/>
  - 吾与点 Wu Yu Dian: <https://www.wuyudian.net/>
- 吾与点平台的实操需调用 LLM API 完成。为确保实操效果，建议提前充值约 **100 ~ 300 RMB** 或 **20,000 ~ 60,000 KRW**。充值路径：[登录] → [充值积分] → [购买积分]，支持 Credit Card / Apple Pay / Google Pay / Alipay / WeChat Pay
- Hands-on exercises on Wu Yu Dian require LLM API calls. To ensure good practice, please

recharge your account in advance with approximately **100–300 RMB** or **20,000–60,000 KRW**. Top-up path: [Login] → [Top Up Credits] → [Purchase Credits]. Supported methods: Credit Card / Apple Pay / Google Pay / Alipay / WeChat Pay.

## 5.2 个人研究材料（推荐）      **Personal Research Materials (Recommended)**

本工作坊鼓励学员以**自己的研究材料**为主线，在课堂上完整体验从古籍数字化、智能表格构建到智能体搭建的全流程，最终带走一个真正属于自己的智能研究助手。若您暂未准备个人材料也无妨：授课团队将提供示范案例供学员跟练，同样可以掌握全流程方法。

This workshop encourages participants to use their **own research materials** as the central thread, so that by the end you take home a research assistant truly built around your own work. If you have not prepared personal materials, that is fine — the teaching team will provide demonstration cases for hands-on practice, and you will still master the complete workflow.

如 希 望 使 用 个 人 材 料 ， 请 按 以 下 要 求 准 备 ：  
If you wish to use your own materials, please prepare them as follows:

- **字数 Word count:** 约 1 万字（最多不超过 2 万字） / approximately 10,000 words (max 20,000)
- **格式 Format:** 支持多种模态，包括 Word / txt / PDF / 图像 (JPG / PNG)；古籍材料推荐使用清晰的图像或 PDF 扫描件 / multimodal supported (Word / txt / PDF / JPG / PNG); for classical texts, clear images or PDF scans are recommended
- **语言 Language:** 中文（古汉语 / 现代汉语）、韩语、英语、日语等均支持 / Chinese (classical / modern), Korean, English, Japanese, etc. all supported
- **要求 Requirement:** 选择自己熟悉、最好已有初步研究问题的材料；材料开头附 **1–2 句研究问题描述**，便于授课团队预检并给出字段设计建议 / choose materials you are familiar with and ideally have a preliminary research question for; please prepend **1–2 sentences describing your research question** so the team can pre-review and suggest field designs
- **提交时间 Submission:** 开课 **3 天** 提交至授课团队邮箱 [gdhc@pku.edu.cn](mailto:gdhc@pku.edu.cn)，邮件标题为【吾与点 Workshop 个人材料】 / submit to [gdhc@pku.edu.cn](mailto:gdhc@pku.edu.cn) at least **3 days before the workshop**, with the email subject line "**Wu Yu Dian Workshop — Personal Materials**"

---

本大纲由北京大学数字人文研究中心编制

This syllabus is prepared by the Research Center for Digital Humanities of Peking University.

联系方式 Contact: [gdhc@pku.edu.cn](mailto:gdhc@pku.edu.cn)

官方网站 Website: <https://pkudh.org/>