

Sessions (/sessions) > Workshop

Preparing Data for LLM Applications Using Data Prep Kit

Data |

When building data intensive applications, a significant portion of your time will be dedicated to data wrangling (cleaning, de-duping, removing markups, etc.). Data Prep Kit (<https://github.com/IBM/data-prep-kit>) is a new open source project that helps you with this.

Data Prep Kit (DPK) is an open source python library that can scale from your laptop to a highly scalable cluster in the cloud. It has been used at scale to prepare terabytes of data to train the IBM Granite Large Language Models (LLMs).

A few noteworthy features of DPK include: de-duping documents (exact dedupe and fuzzy dedupe), handling documents and code, language detection (spoken languages and programming languages), removing PII, malware detection and creating embeddings for a vector database.

In this talk, I will go over some interesting features of the Data Prep Kit. If time permits, I will show a demo.

Event Details

Time & Duration:

11:45 am

90 – 120 min

In-Person Location:

Lovelace Room

Hi, Welcome to Ai Summit

How can I help you today?

Watch online:

Watch live broadcast on Zoom (<https://us06web.zoom.us/j/88251412080?pwd=VITnOXqdOnarsNltzIBFBizNI6NBSa.1>).



Featured Speaker



(<https://techequity-ai.org/speaker/sujee-maniyam/>)

Sujee Maniyam

AI Engineer

Node51

Sujee Maniyam is a seasoned practitioner focusing on Generative AI, Machine Learning, Deep Learning, Big Data, Distributed Systems, and Cloud technologies. He also love teaching and has taught and mentored thousands of professionals.

+ Learn more

(<https://techequity-ai.org/speaker/sujee-maniyam/>)

Other Speakers Sessions

Hi, Welcome to Ai Summit
How can I help you today?

Collaborate To Innovate And Advance.

Begin your journey of collaborative innovation and collective growth with Ai Summit. Your participation helps propel the democratization of AI knowledge and applications.

Register Now (/registration/)

Get Involved

Career(/career-overview)

Become a Speaker(/become-a-speaker)

Become a Volunteer(/career-overview)

About Us

Who We Are(/about-us/)

Why Attend?(/about-us/)

ai.4.all Program(/frequently-asked-questions/)

Resources

Blog(/blog/)

Event Recordings(/)

Support

Contact Us(/contact-us/)

FAQs(/frequently-asked-questions/)

Privacy Policy(/privacy-policy)

Terms of Use(/)

Ai Summit


Fall 2024 (/ai-summit-spring-2024/)


Agenda(/agenda-fall-2024/)

Speakers(/speakers-fall-2024/)

Sessions(/sessions-fall-2024/)

© 2024 Ai Summit. All rights reserved by TechEquity Inc., a 501(c)(3) nonprofit organization.

 (<https://www.linkedin.com/company/ai-summit-org/>)

 (<https://www.facebook.com/ai.summit.org>)

 (https://www.instagram.com/ai.summit_org/) Hi, Welcome to Ai Summit
How can I help you today?

 (<https://www.youtube.com/@AiSummitSiliconValley>)

 (<http://www.meetup.com/ai-summit>)  (https://ai_summit_org.eventbrite.com)