Operating System (OS) in the Era of AI

Xiao-Feng Li
xiaofeng.li@gmail.com
May 20, 2023
Evolution of Computing: User Interactions

Text and command  Mouse and GUI  Touch and mobile  Natural interaction

**Transition**

*Users serve computer ➔ Computers serve users*
Core in Computing: Operating System

Role of Operating System

*Bridge user tasks and hardware (such as Windows, iOS, Android, Linux, etc.)*
Evolution of Computing: Functionalities

Text and command
Mouse and GUI
Touch and mobile
Natural interaction

Task and disk
Gfx and browsing
Media and comm.
Perception and prediction
Core in Computing: Current Status

Design goal of Operating System

hardware resource utilization → user value and experience

Apps
Operating system
Hardware

Details

Intention Sensing
(gesture/voice/vision/location/context)

Immersive Presenting
(graphic/audio/video/AR/XR)

Compute
(CPU/GPU/TPU)

Data
(scalar/vector/tensor)

Wireless
(NFC/Bluetooth/WiFi/5G)
Evolution of OS Stack

Thin client becomes thicker
Challenges to AI on End Devices

AI model size increases dramatically faster than chip capability

#Parameters:
GPT3.5: 175 billion
GPT4.0: ~ 1 trillion

Source: MIT INITIATIVE ON THE DIGITAL ECONOMY RESEARCH BRIEF 2020 Vol. 4
Evolution of OS Stack

Thin client becomes thicker and thinker
Opportunities of Next-Gen OS

- Understanding user needs
  *Low power always-on sensing and end-to-end trusted computing*
- Execution of AI tasks
  *AI model inference acceleration and heterogenous computing*
- Intuitive user interaction
  *Multi-modal interaction and smart in-context learning*
Thanks!