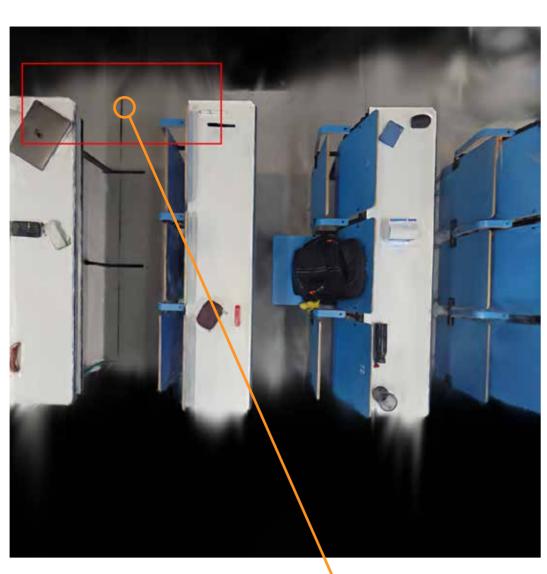Prompt: The calculator in the first table is dirty, please bring some tissues.

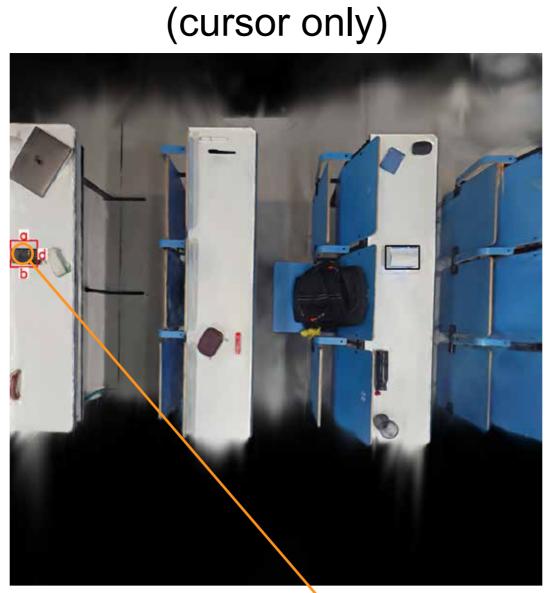GPT-4o — Coordinate: [200,200].

Visual CoT — Coordinate: [252,173].
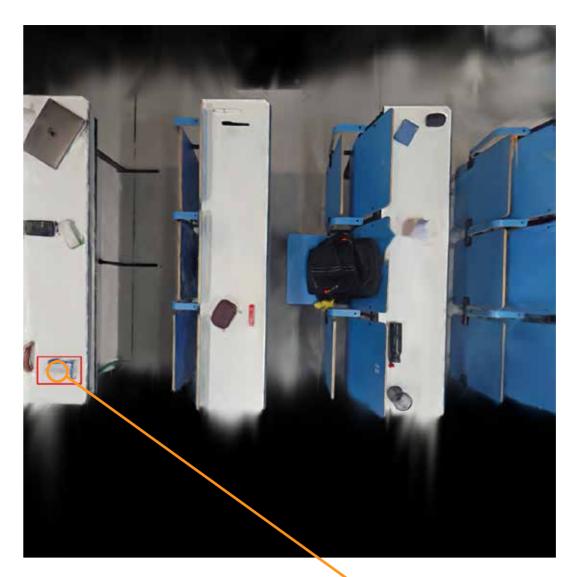
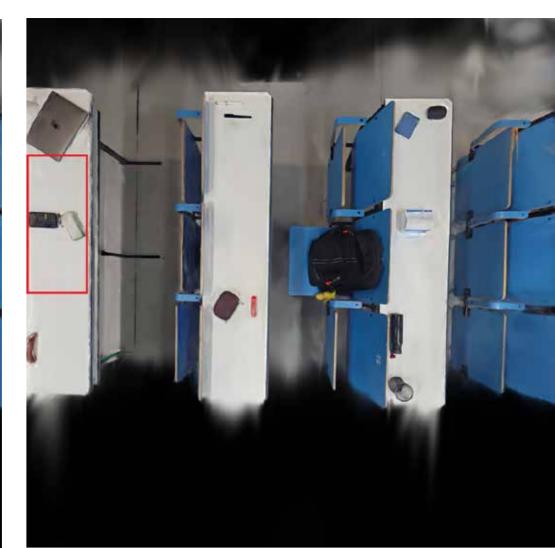Autonomous Imagination (cursor only) — Coordinate: [31,451].
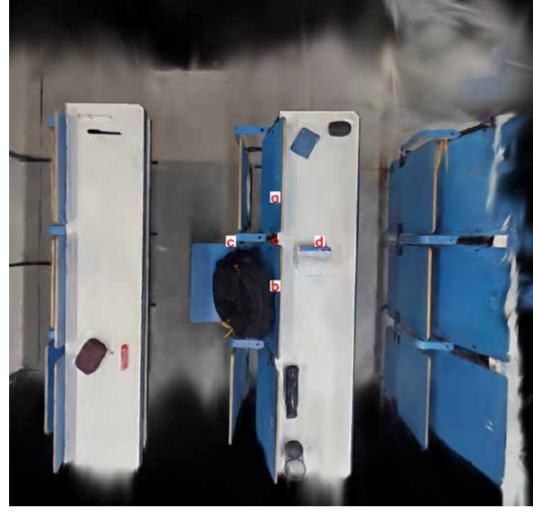
Sampling — Coordinate: [80,815].

Ground Truth — Coordinate: [72,417].

Autonomous Imagination (ours)
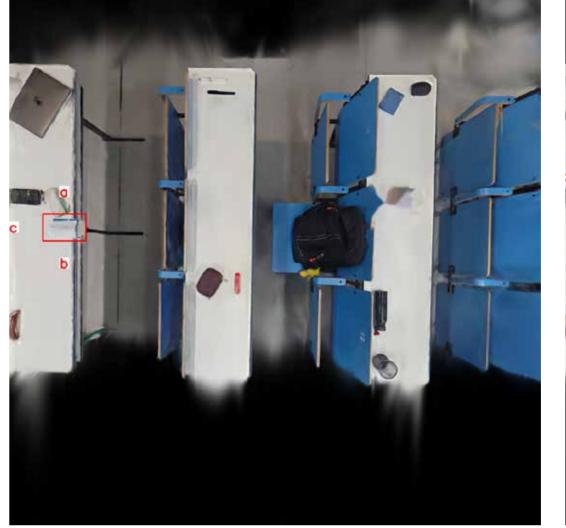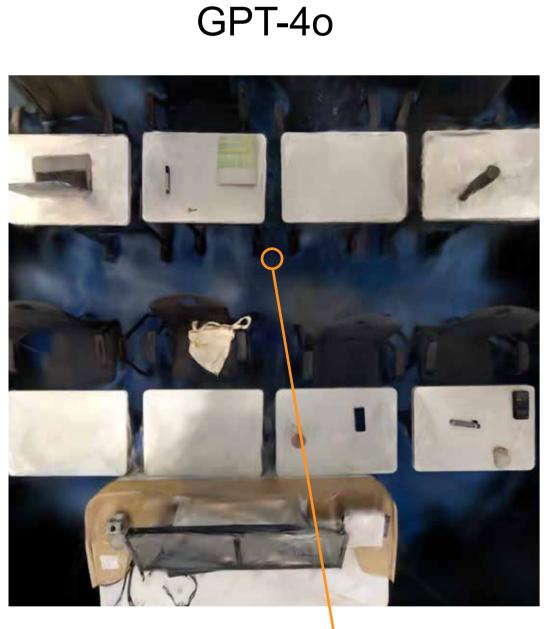
A student is sitting in a seat with a white bag on it. Could you place the black cell phone on her table?
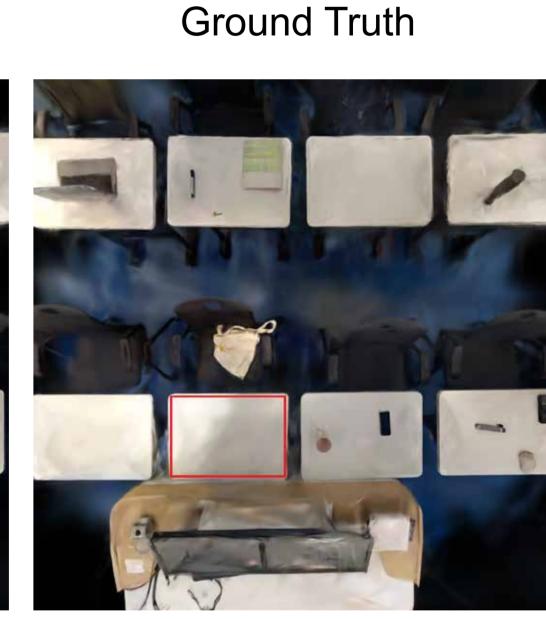
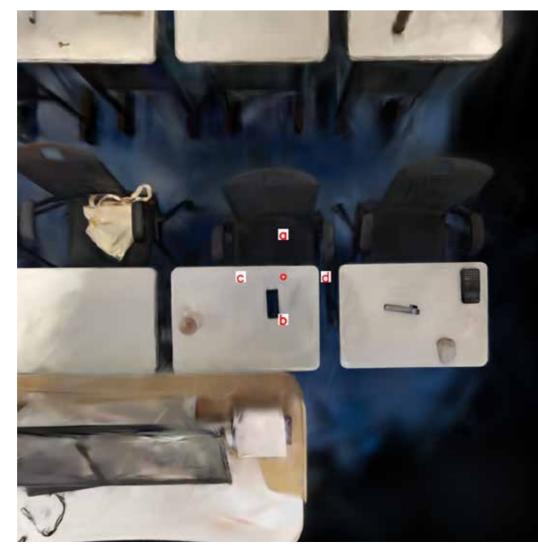GPT-4o — Coordinate: [600,400].

Visual CoT — Coordinate: [490,1037].

Autonomous Imagination (cursor only) — Coordinate: [493,576].

Sampling — Coordinate: [450,759].

Ground Truth — Coordinate: [397,754].

Autonomous Imagination (ours)