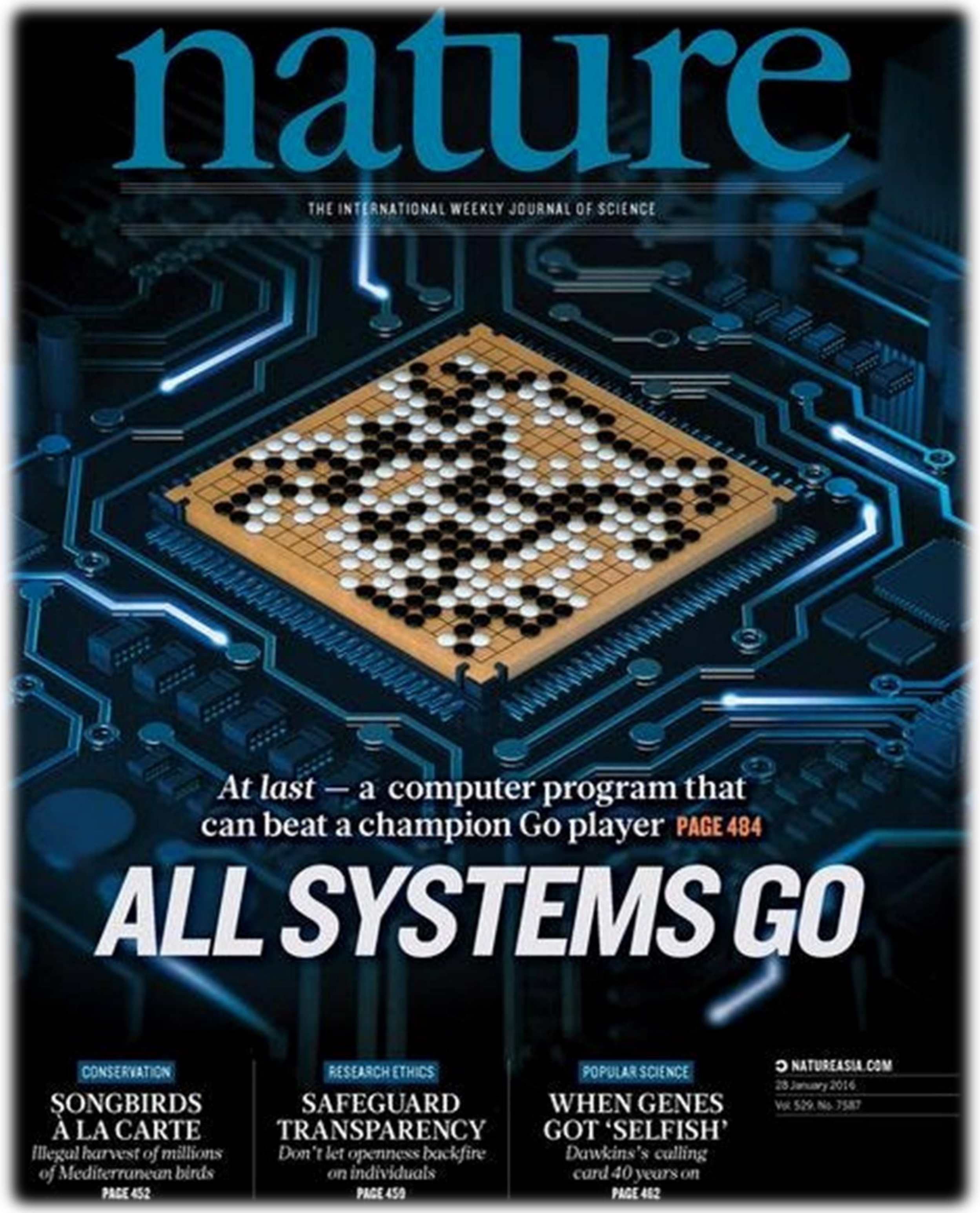
Towards Environmentally *Sustainable* & *Equitable* Computing

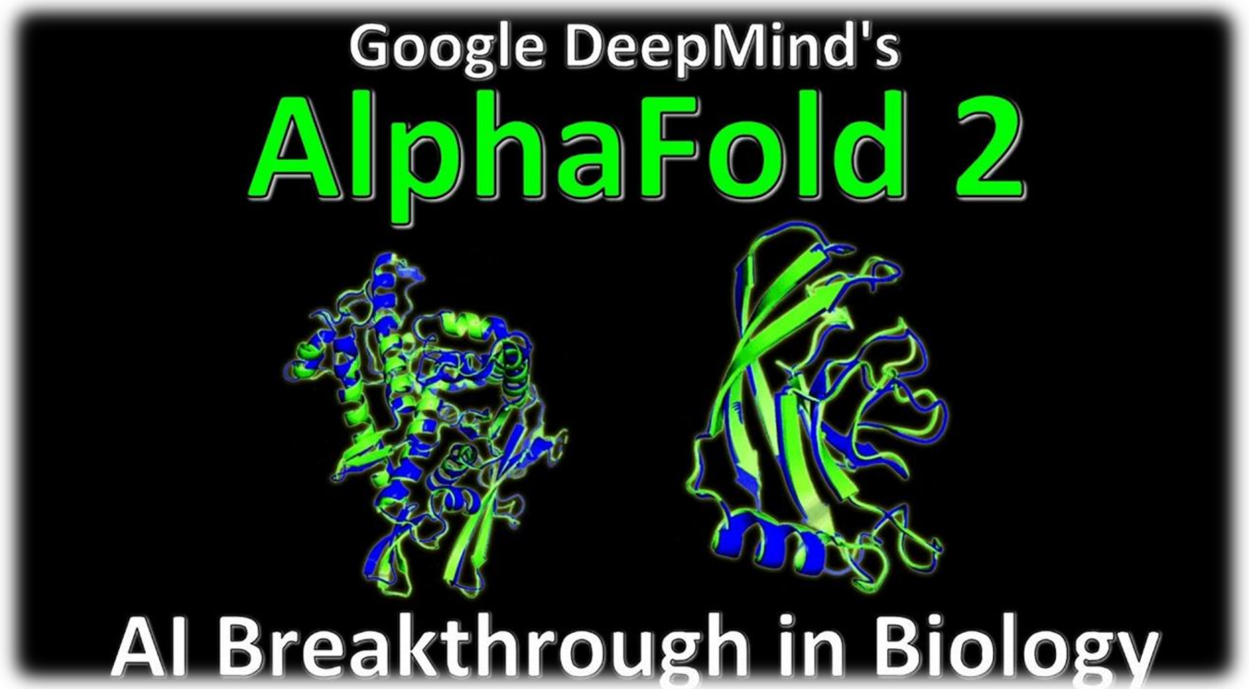
Shaolei Ren

UC RIVERSIDE

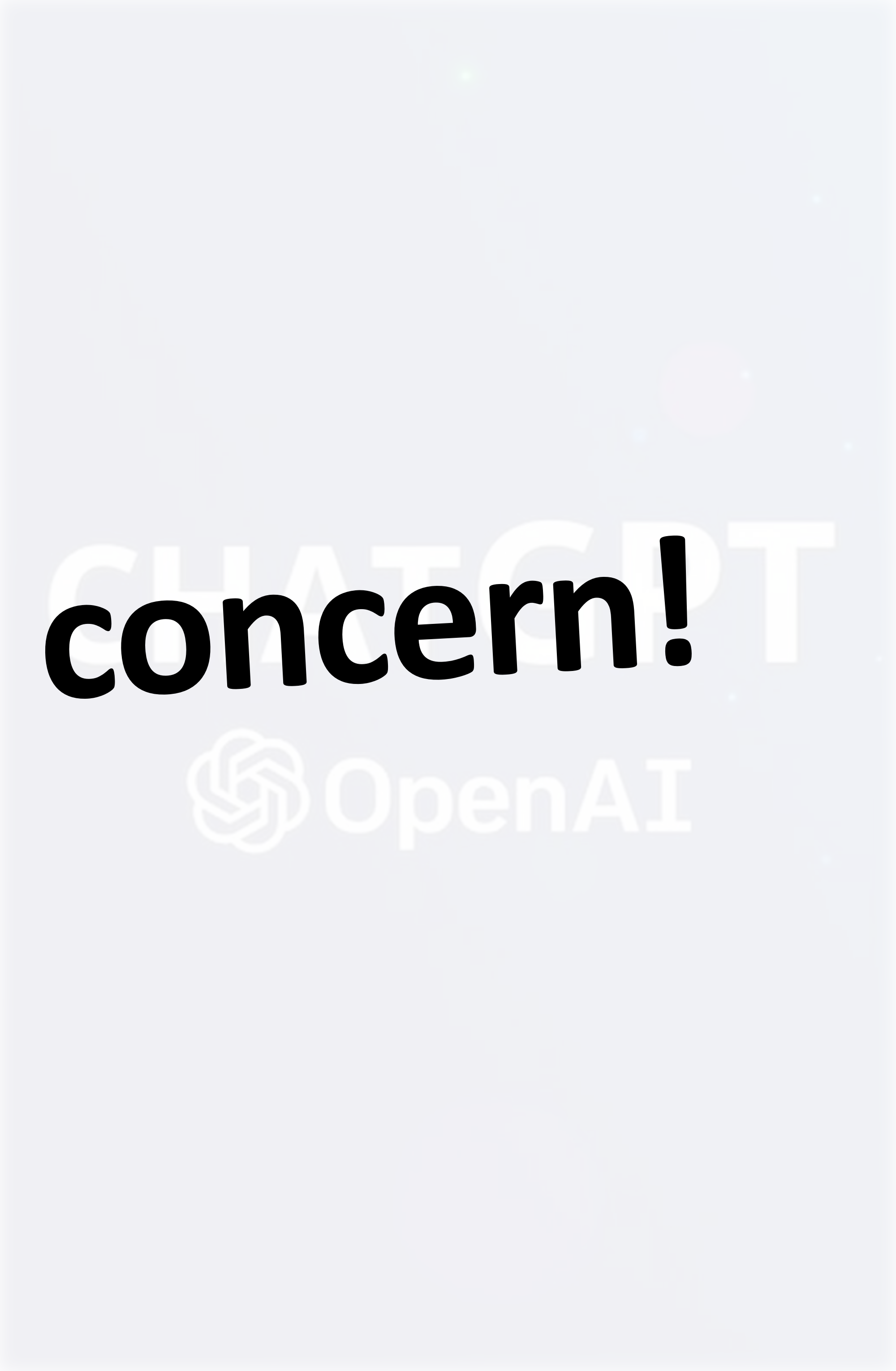
Partly based on joint work with Pengfei Li (UCR), Jianyi Yang (UCR), Mohammad A. Islam (UT Arlington) and Adam Wierman (Caltech).

The demand for AI and computing is soaring!





The demand for AI and computing is soaring!



… and so is the sustainability concern!

Computing is resource-intensive and power-hungry

Computing is resource-intensive and power-hungry

Computing is resource-intensive and power-hungry

Per IEA, data centers use about 300 TWh electricity in 2022, or 1-2% of the global electricity demand (0.2-0.4% global energy demand).

Electricity = Carbon

Electricity = Carbon + Water + Air & thermal pollution + …

Electricity = Carbon + Water + Air & thermal pollution + ...

Carbon != Water

One footprint is "*a complement to and not a substitute for*" the other.

Kai Fang, Reinout Heijungs, Geert R. de Snoo, Theoretical exploration for the combination of the ecological, energy, carbon, and water footprints: Overview of a footprint family, Ecological Indicators, Volume 36, 2014,

Water withdrawal vs. water consumption for electricity

Take water from a source (e.g., groundwater)

Water "lost" (due to, e.g., evaporation) and not returning to the source

Water withdrawal vs. water consumption for electricity

Take water from a source (e.g., groundwater)

Water "lost" (due to, e.g., evaporation) and not returning to the source

1 kWh = roughly 44-100 L water withdrawal (excluding hydropower)

1 kWh = 3.14 L water consumption

Based on U.S. average water efficiency for electricity generation provided by EIA and WRI.

Data centers are guzzling water!

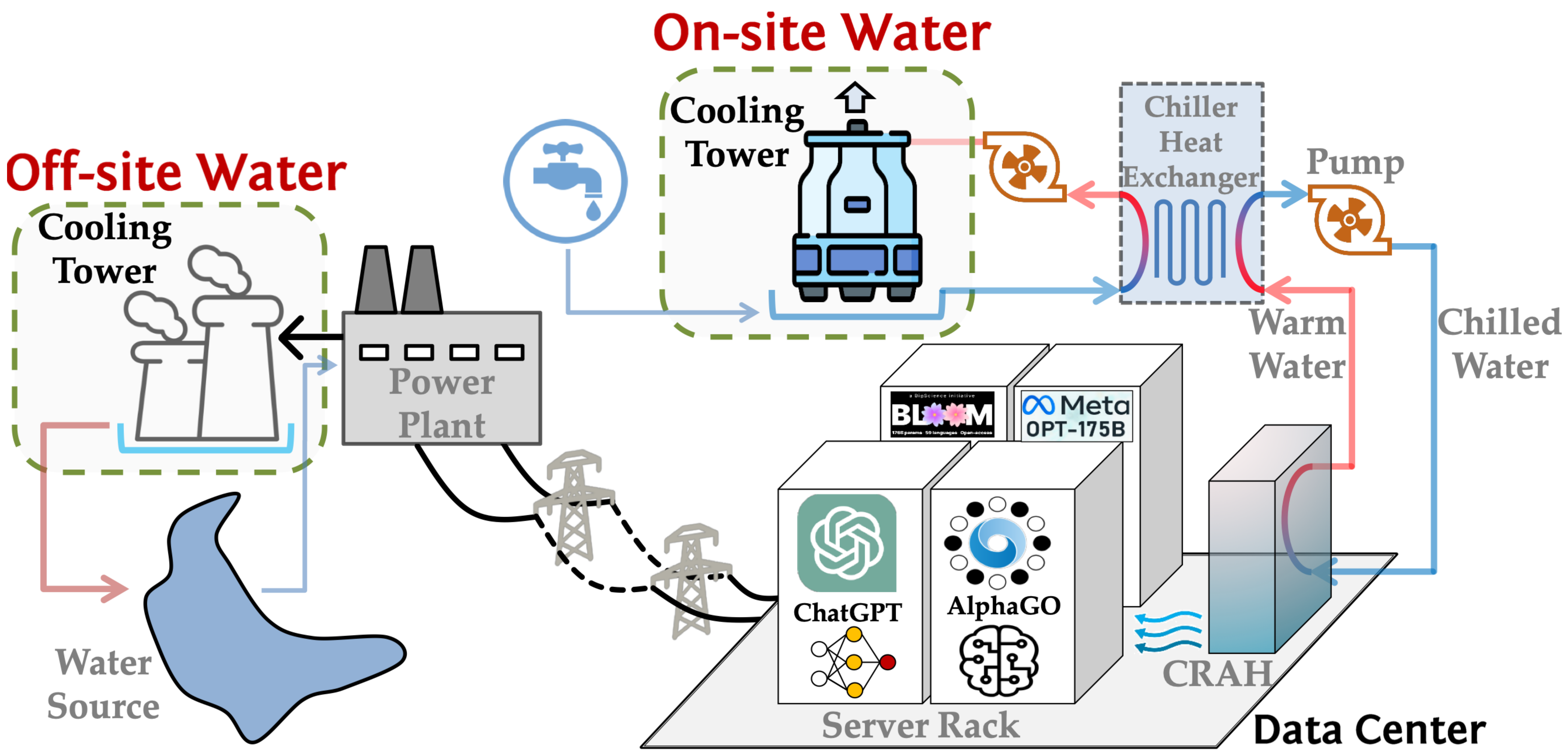
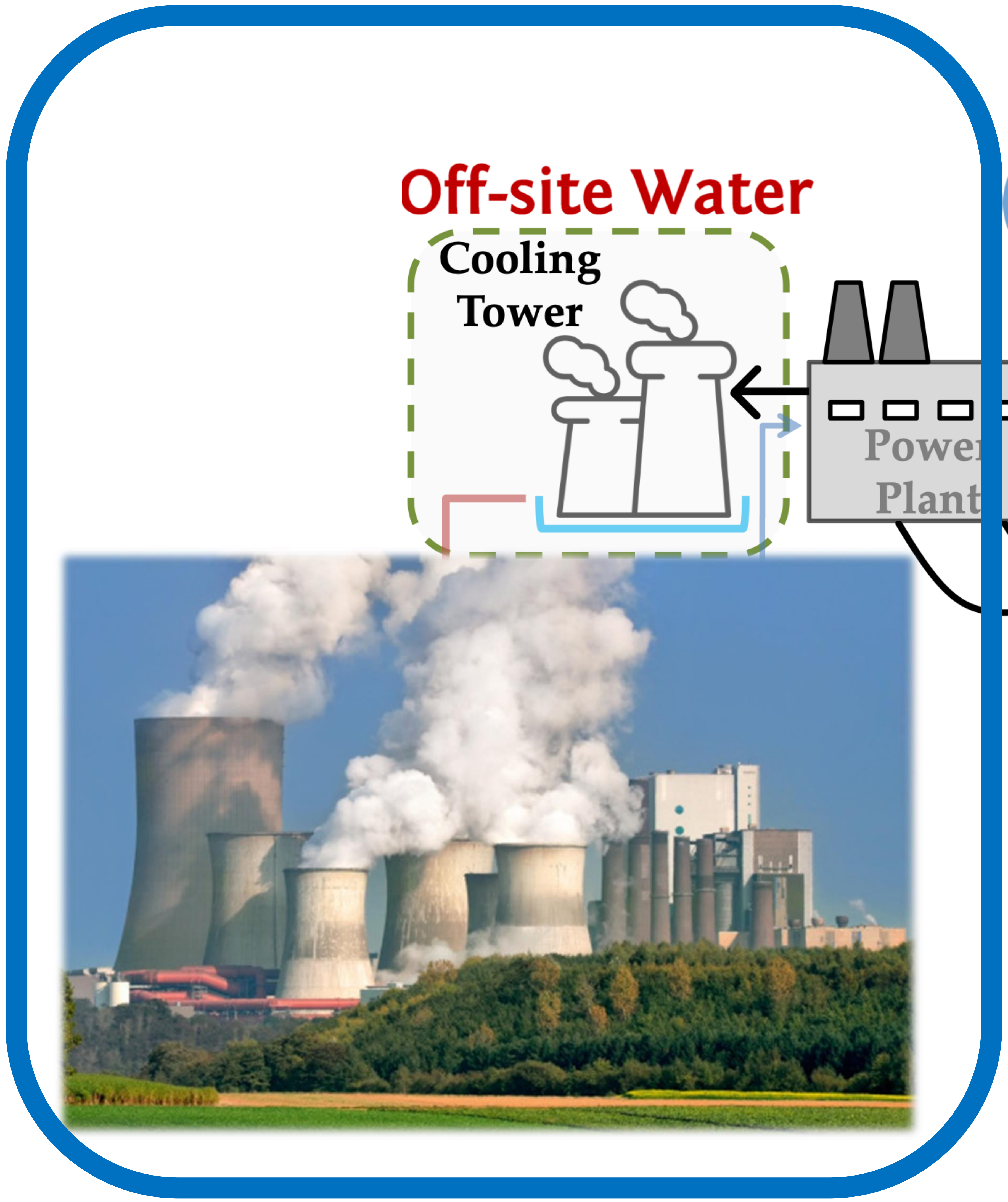


Figure 2. Multiple AI models are trained and/or deployed in the data center. Data center water footprint consists of two parts: on-site water and off-site water consumption

Data centers are guzzling water!



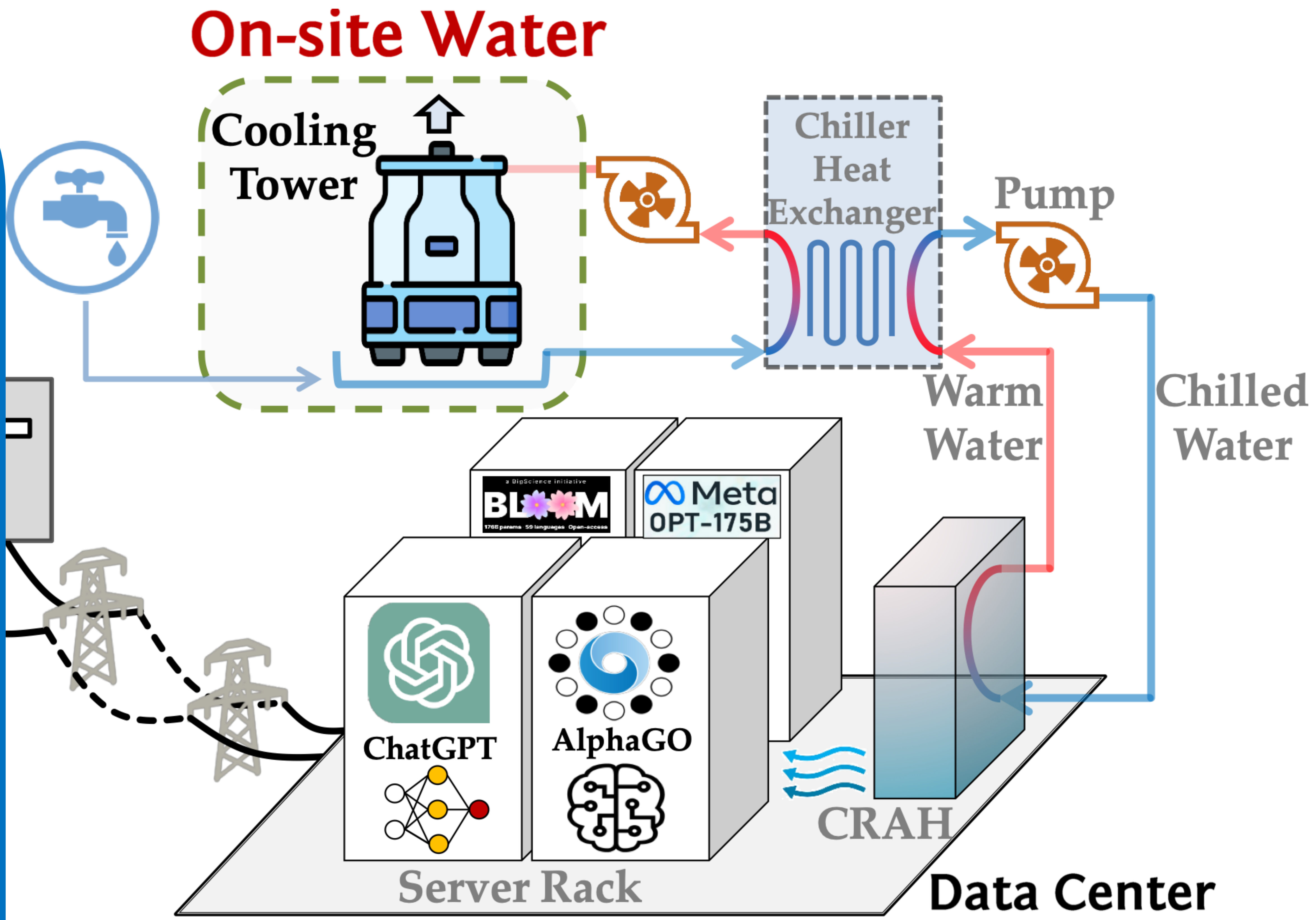


Figure 2. Multiple AI models are trained and/or deployed in the data center. Data center water footprint consists of two parts: on-site water and off-site water consumption

Data centers are guzzling water!

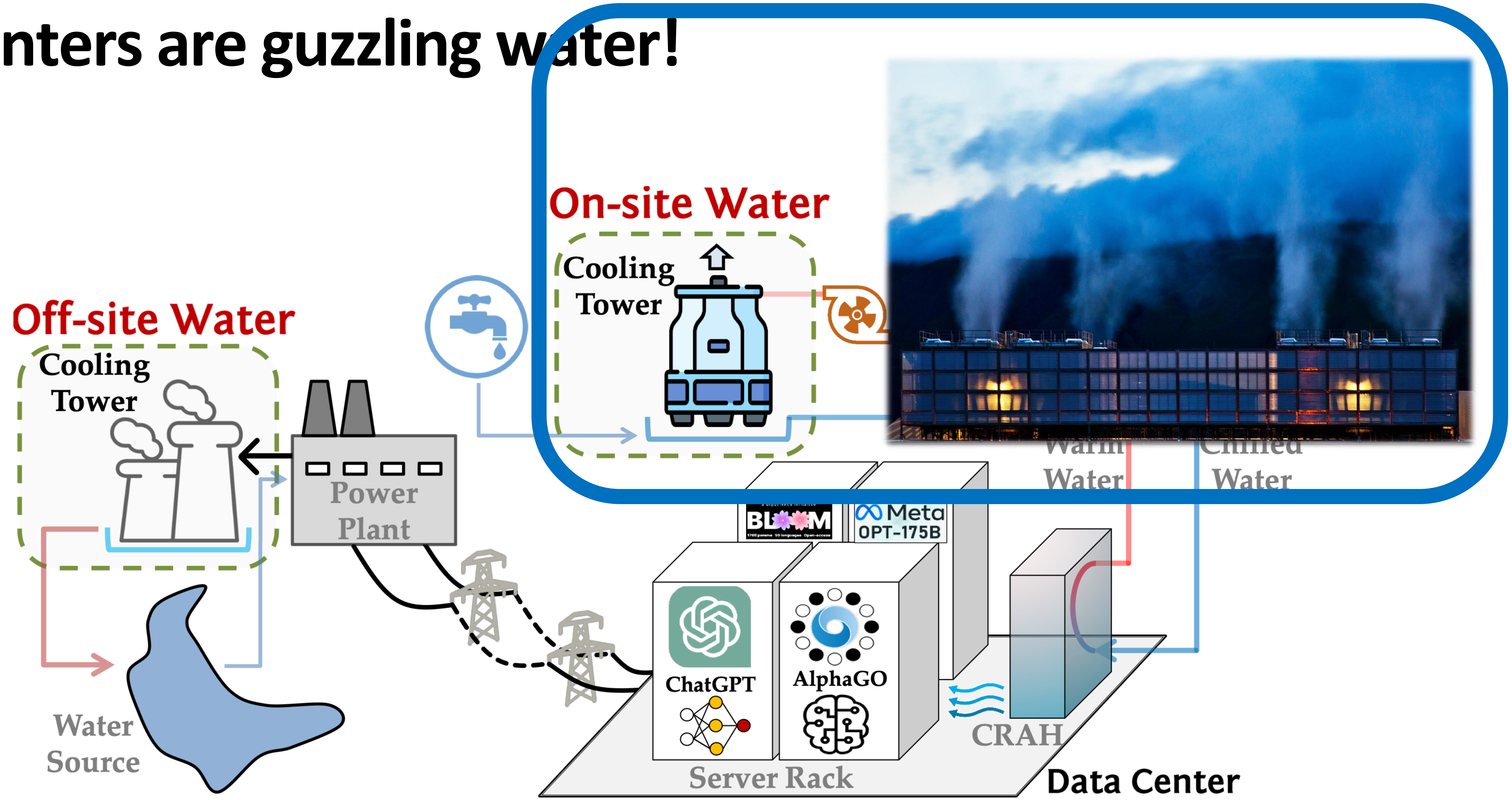
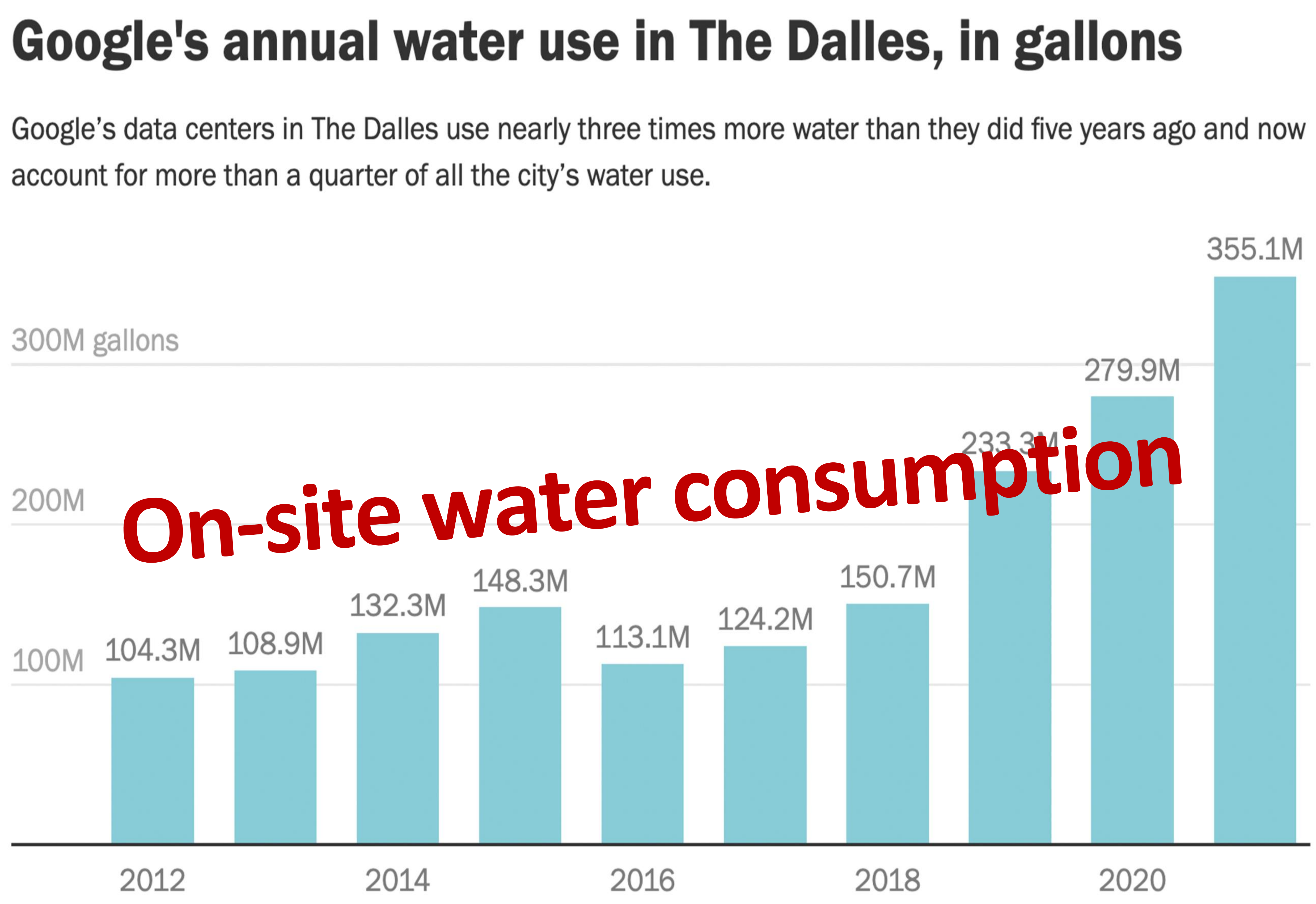


Figure 2. Multiple AI models are trained and/or deployed in the data center. Data center water footprint consists of two parts: on-site water and off-site water consumption

Data centers are guzzling water!

- A large data center can consume millions of gallons of potable water each day for on-site cooling.
- Google’s data center used 355 million gallons of water in The Dalles, OR, in 2021, 29% of the city’s total water consumption



And this trend continues!

What about AI's water usage (withdrawal)?

~6.6 billion cubic meters in 2027

Based on the projected AI GPU energy consumption of up to 134 TWh (de Vries, 2023), US average water withdrawal for electricity generation 44L/kWh, PUE=1.1, on-site water withdrawal 1L/kWh.

What about AI's water usage (consumption)?

~0.6 billion cubic meters in 2027

Based on the projected AI GPU energy consumption of up to 134 TWh (de Vries, 2023), US average water consumption for electricity generation 3.14 L/kWh, PUE=1.1, on-site water consumption 0.8L/kWh.

ChatGPT is already "drinking" a lot of water

Table 2: *Estimate of GPT-3's average operational water footprint. "*" denotes data centers under construction as of July 2023, and the PUE and WUE values for these data centers are based on Microsoft's projection.*

Location	**PUE**	**WUE** (L/kWh)	**Electricity Water Intensity** (L/kWh)	**Water for Training** (million L)			**Water for Inference** (mL)			**# of Inferences for 500ml Water**
				Onsite Water	Offsite Water	Total Water	Onsite Water	Offsite Water	Total Water	
US Average	1.170	0.550	3.142	0.708	4.731	**5.439**	2.200	14.704	**16.904**	**29.6**
Wyoming	1.125	0.230	2.574	0.296	3.727	**4.023**	0.920	11.583	**12.503**	**40.0**
Iowa	1.160	0.190	3.104	0.245	4.634	**4.879**	0.760	14.403	**15.163**	**33.0**
Arizona	1.223	2.240	4.959	2.883	7.805	**10.688**	8.960	24.259	**33.219**	**15.1**
Washington	1.156	1.090	9.501	1.403	14.136	**15.539**	4.360	43.934	**48.294**	**10.4**
Virginia	1.144	0.170	2.385	0.219	3.511	**3.730**	0.680	10.913	**11.593**	**43.1**
Texas	1.307	1.820	1.287	2.342	2.165	**4.507**	7.280	6.729	**14.009**	**35.7**
Singapore	1.358	2.060	1.199	2.651	2.096	**4.747**	8.240	6.513	**14.753**	**33.9**
Ireland	1.197	0.030	1.476	0.039	2.274	**2.313**	0.120	7.069	**7.189**	**69.6**
Netherlands	1.158	0.080	3.445	0.103	5.134	**5.237**	0.320	15.956	**16.276**	**30.7**
Sweden	1.172	0.160	6.019	0.206	9.079	**9.284**	0.640	28.216	**28.856**	**17.3**
Mexico*	1.120	0.056	5.300	0.072	7.639	**7.711**	0.224	23.742	**23.966**	**20.9**
Georgia*	1.120	0.060	2.309	0.077	3.328	**3.406**	0.240	10.345	**10.585**	**47.2**
Taiwan*	1.200	1.000	2.177	1.287	3.362	**4.649**	4.000	10.448	**14.448**	**34.6**
Australia*	1.120	0.012	4.259	0.015	6.138	**6.154**	0.048	19.078	**19.126**	**26.1**
India*	1.430	0.000	3.445	0.000	6.340	**6.340**	0.000	19.704	**19.704**	**25.4**
Indonesia*	1.320	1.900	2.271	2.445	3.858	**6.304**	7.600	11.992	**19.592**	**25.5**
Denmark*	1.160	0.010	3.180	0.013	4.747	**4.760**	0.040	14.754	**14.794**	**33.8**
Finland*	1.120	0.010	4.542	0.013	6.548	**6.561**	0.040	20.350	**20.390**	**24.5**

Estimates updated as of 09/2023 based on Microsoft's regional WUE data.

ChatGPT is already "drinking" a lot of water

ChatGPT needs about 500 ml of water for answering 10-50 questions.

Estimates updated as of 09/2023 based on Microsoft's regional WUE data.

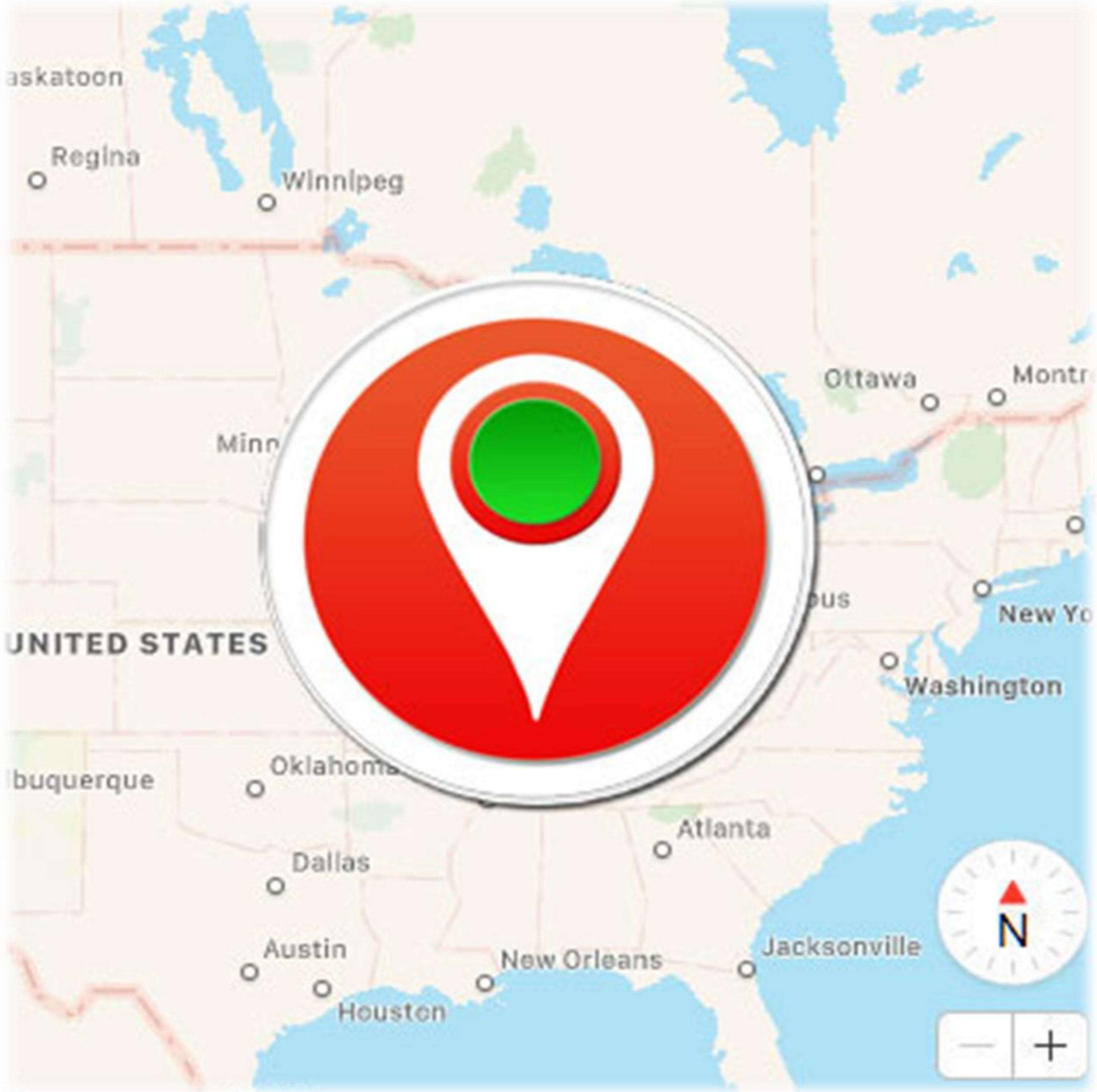
Water is a shared public good…

"Every drop matters." **--- Meta**

Water Positive by 2030!

Algorithmic challenges

“When” and “Where” matter a lot

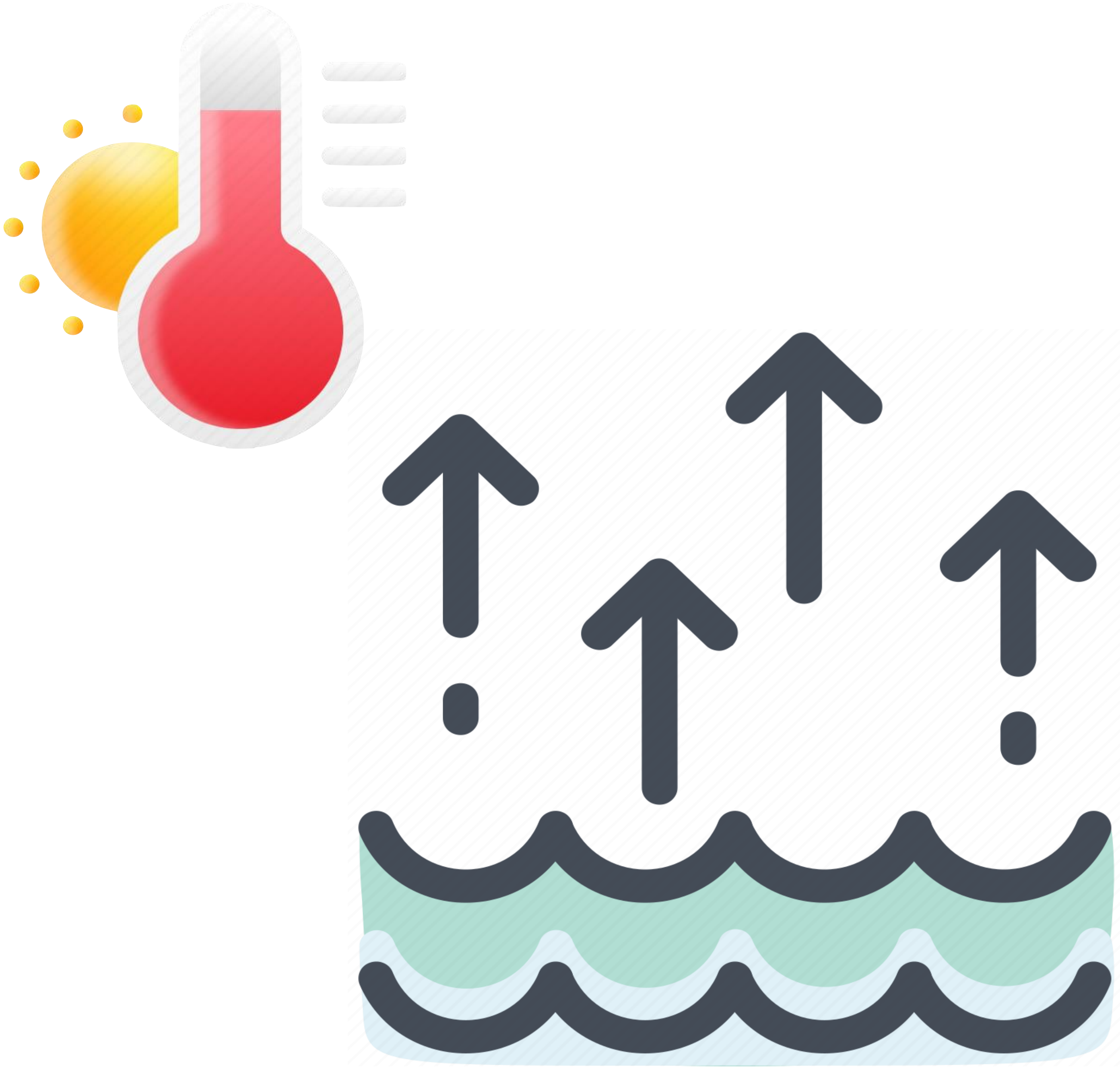


- Outside weather condition
- Time-varying workloads

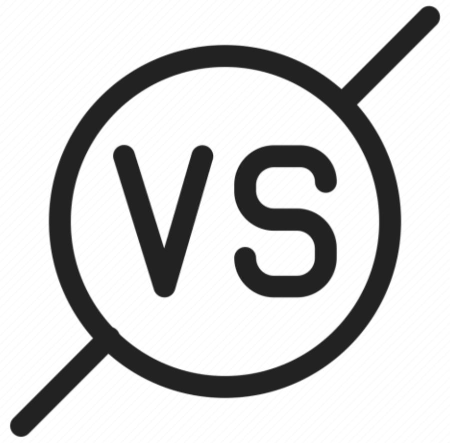
- Fuel mix for power generation
- Regional climate

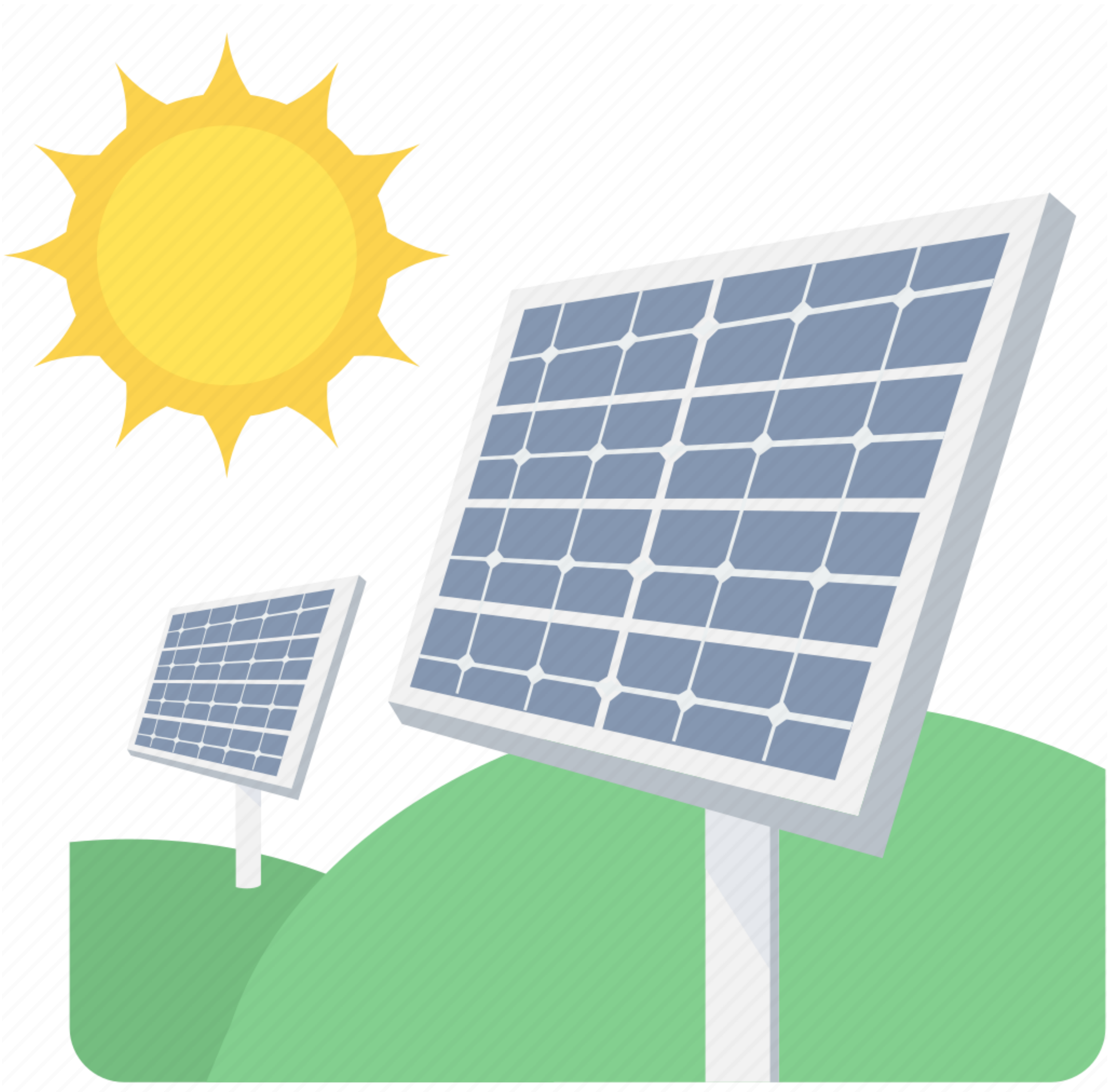
Algorithmic challenges

“Follow the Sun” or “Unfollow the Sun”?



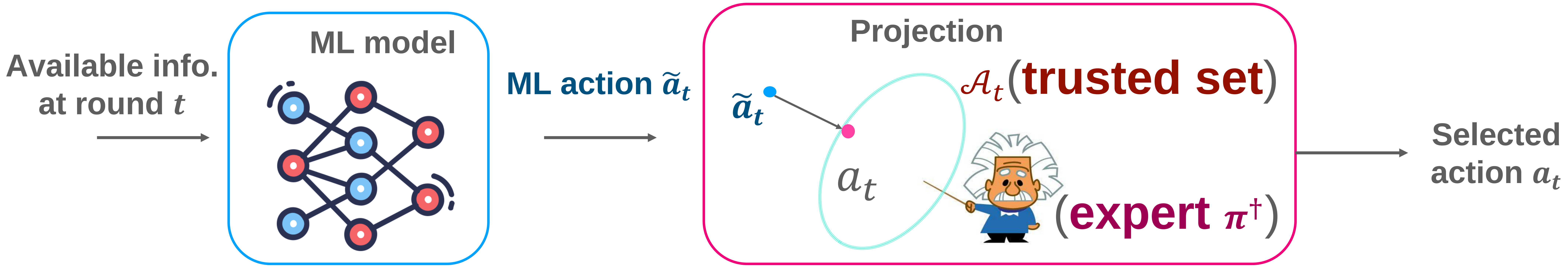
Water Efficient





Carbon Efficient

Learning-augmented algorithms



- **[NeurIPS'23]** Jianyi Yang, Pengfei Li, Tongxin Li, Adam Wierman, and Shaolei Ren, **"Anytime-Constrained Reinforcement Learning with Policy Prior,"** NeurIPS, 2023.
- **[NeurIPS'23]** Pengfei Li, Jianyi Yang, Adam Wierman, and Shaolei Ren, **"Robust Learning for Smoothed Online Convex Optimization with Feedback Delay,"** NeurIPS, 2023.
- **[ICML'23]** Pengfei Li, Jianyi Yang, and Shaolei Ren, **"Learning for Edge-Weighted Online Bipartite Matching with Robustness Guarantees,"** ICML, 2023.
- **[SIGMETRICS'22]** Pengfei Li, Jianyi Yang, and Shaolei Ren, **"Expert-Calibrated Learning for Online Optimization with Switching Costs,"** SIGMETRICS, 2022.

Environmental equity?

minimizing the total environmental cost

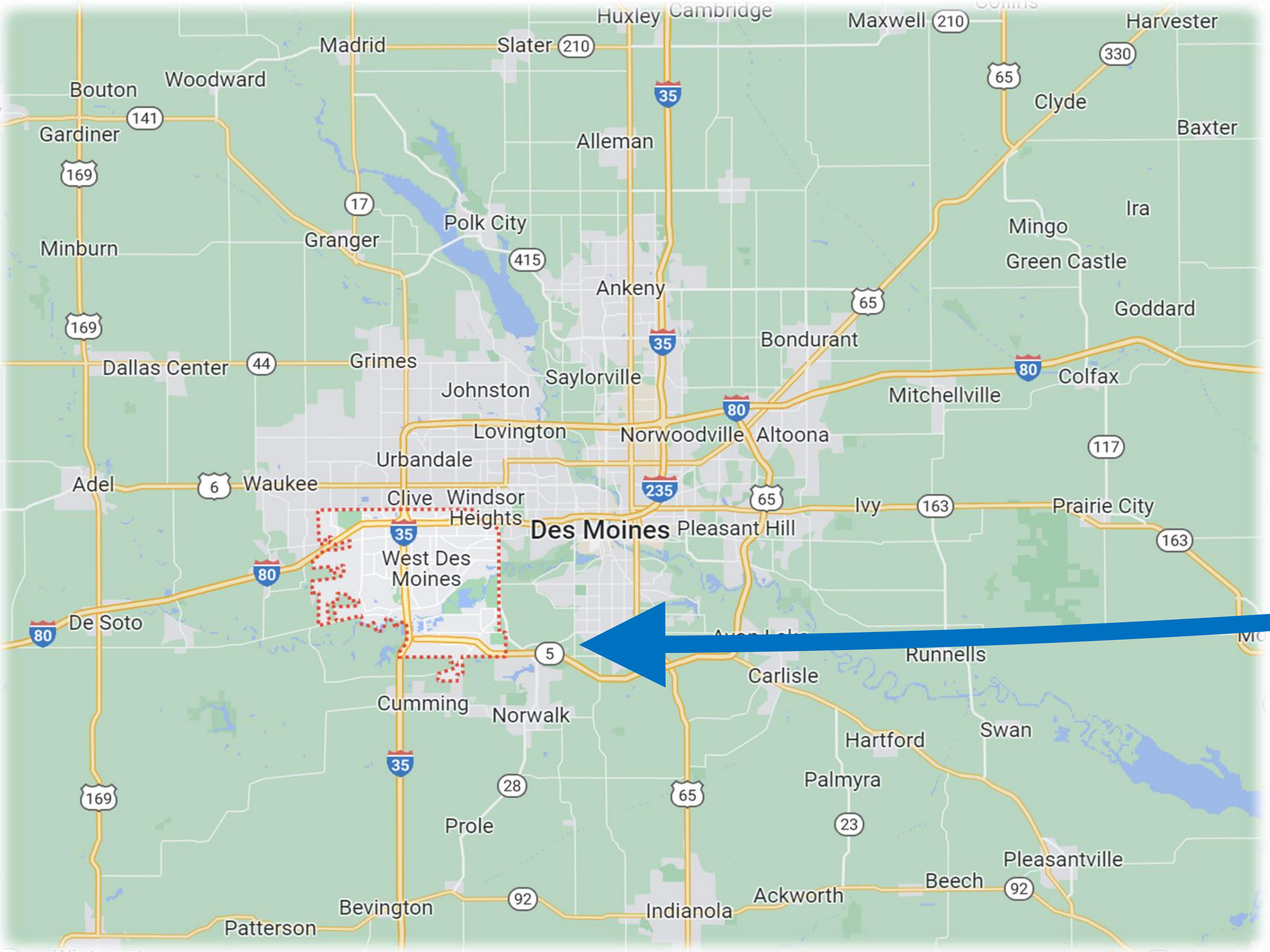
$\neq$

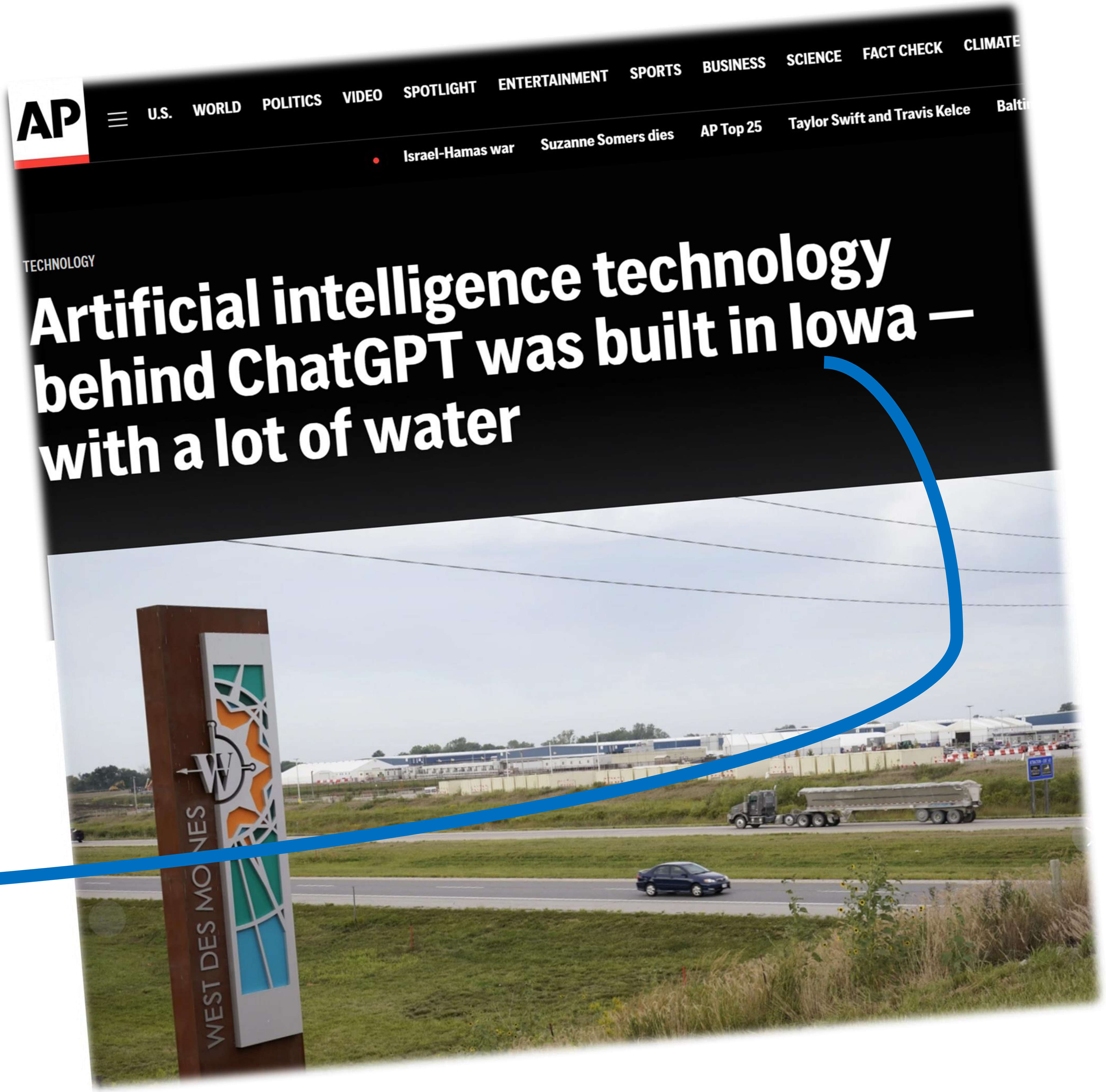
minimizing each region's environmental cost

Electricity = Carbon + Water + Air & thermal pollution + …

Electricity = Carbon + Water + Air & thermal pollution + …

OpenAI
ChatGPT 4.0





Computing's environmental inequity is emerging…

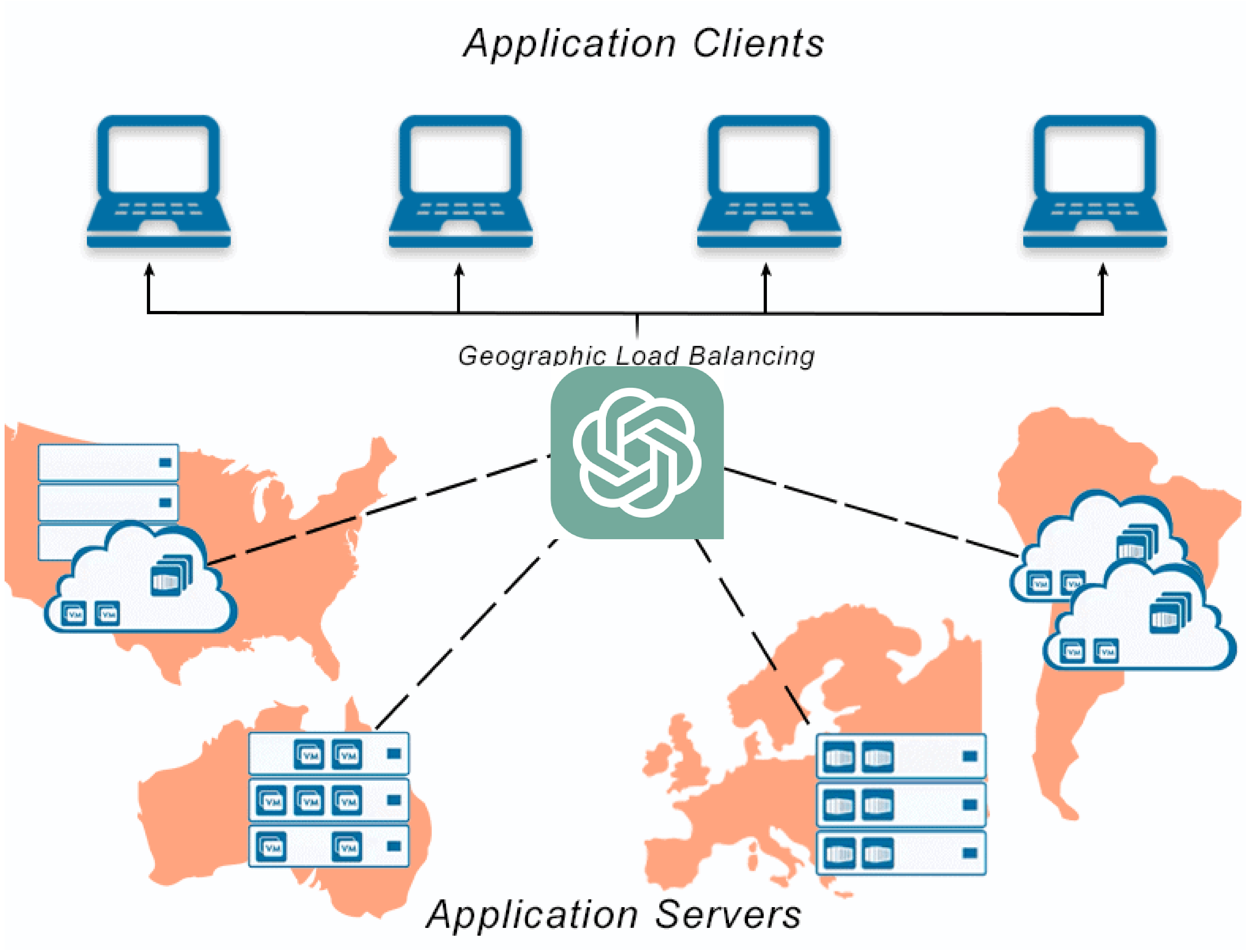
AINOW

The constant push for scale in artificial intelligence has led Big Tech firms to develop hugely energy-intensive computational models that optimize for "accuracy" – through increasingly large datasets and computationally intensive model training – over more efficient and sustainable alternatives. (1) As we increasingly become locked into using Big-Tech infrastructures, we also become locked into their voracious appetite for resources, necessitating a life cycle analysis: the data centers needed for computationally intensive AI have high energy costs and carry a massive carbon footprint. (2) Computing technologies rely heavily on minerals that are procured under violent and exploitative conditions. (3) But these environmental harms are not evenly distributed; they disproportionately impact communities that are already marginalized, in a manner that reenacts historical practices of settler colonialism and racial capitalism. (4)

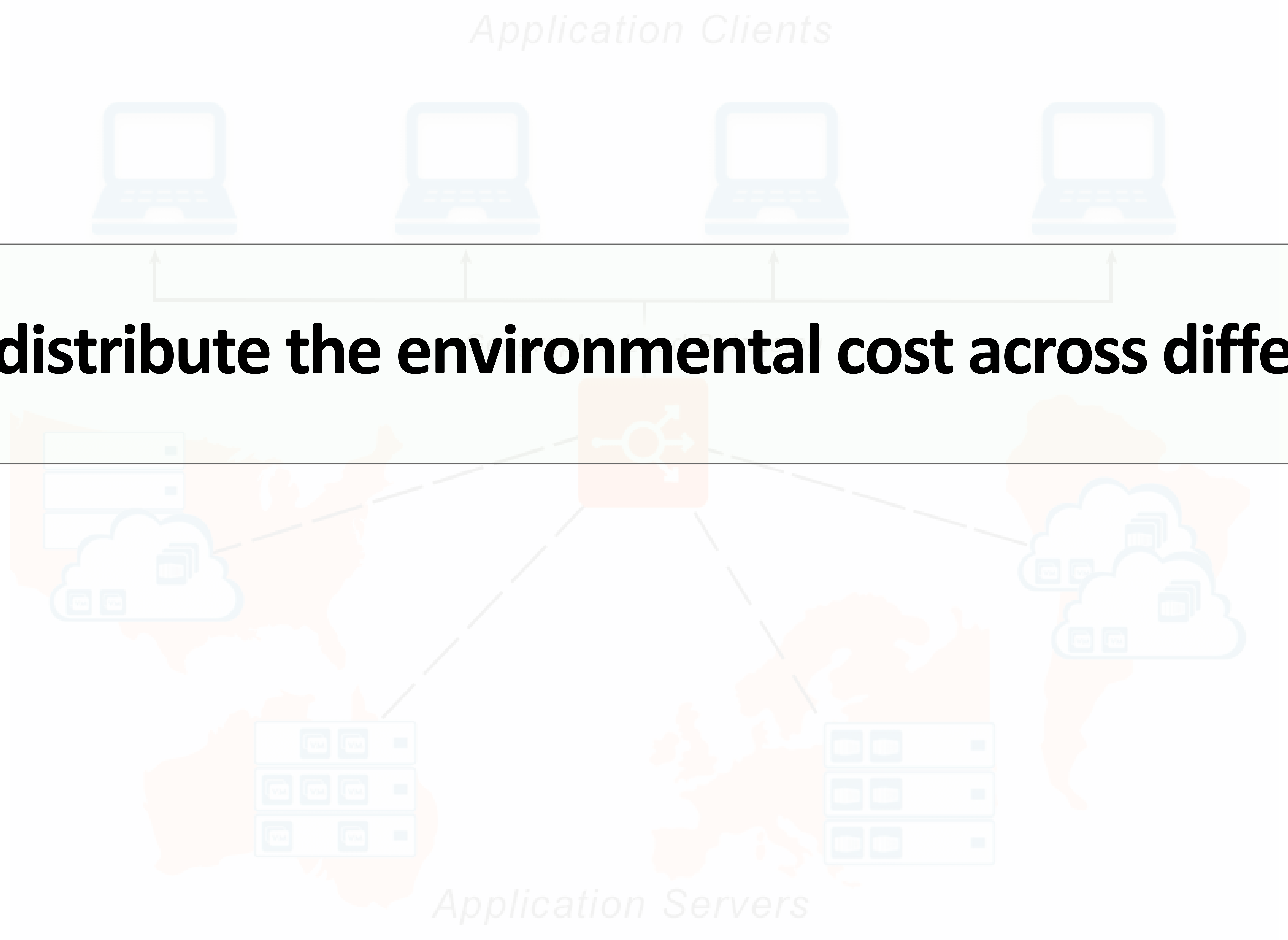
The uneven distribution of AI's environmental cost is "**historical practices of settler colonialism and racial capitalism**".

In 2022, the **United Nations Educational, Scientific and Cultural Organization** (UNESCO) recommends that "AI should not be used" if it creates "disproportionate negative impacts on the environment".

Environmentally *equitable* geographical load balancing (eGLB)



Environmentally *equitable* geographical load balancing (eGLB)



Equitably **re-distribute the environmental cost across different regions!**

Computing is “thirsty”, just as we are!

ChatGPT needs 500 ml of water for answering 10-50 questions.

Estimates updated as of 09/2023 based on Microsoft’s regional WUE data.

The water footprint is coming to the public…

Related papers

- **[NeurIPS'23]** Jianyi Yang, Pengfei Li, Tongxin Li, Adam Wierman, and Shaolei Ren, **"Anytime-Constrained Reinforcement Learning with Policy Prior,"** NeurIPS, 2023.
- **[NeurIPS'23]** Pengfei Li, Jianyi Yang, Adam Wierman, and Shaolei Ren, **"Robust Learning for Smoothed Online Convex Optimization with Feedback Delay,"** NeurIPS, 2023.
- **[ICML'23]** Pengfei Li, Jianyi Yang, and Shaolei Ren, **"Learning for Edge-Weighted Online Bipartite Matching with Robustness Guarantees,"** ICML, 2023.
- **[SIGMETRICS'22]** Pengfei Li, Jianyi Yang, and Shaolei Ren, **"Expert-Calibrated Learning for Online Optimization with Switching Costs,"** SIGMETRICS, 2022.
- [Preprint'23] Pengfei Li, Jianyi Yang, Adam Wierman, and Shaolei Ren, **"Towards Environmentally Equitable AI via Geographical Load Balancing,"** arXiv, 2023.
- [Preprint'23] Pengfei Li, Jianyi Yang, Mohammad A. Islam, Shaolei Ren, **"Making AI Less 'Thirsty': Uncovering and Addressing the Secret Water Footprint of AI Models,"** arXiv, 2023.

Thanks!