Adaptive Data Analysis with Correlated Observations

Aryeh Kontorovich, Meni Sadigurschi & Uri Stemmer

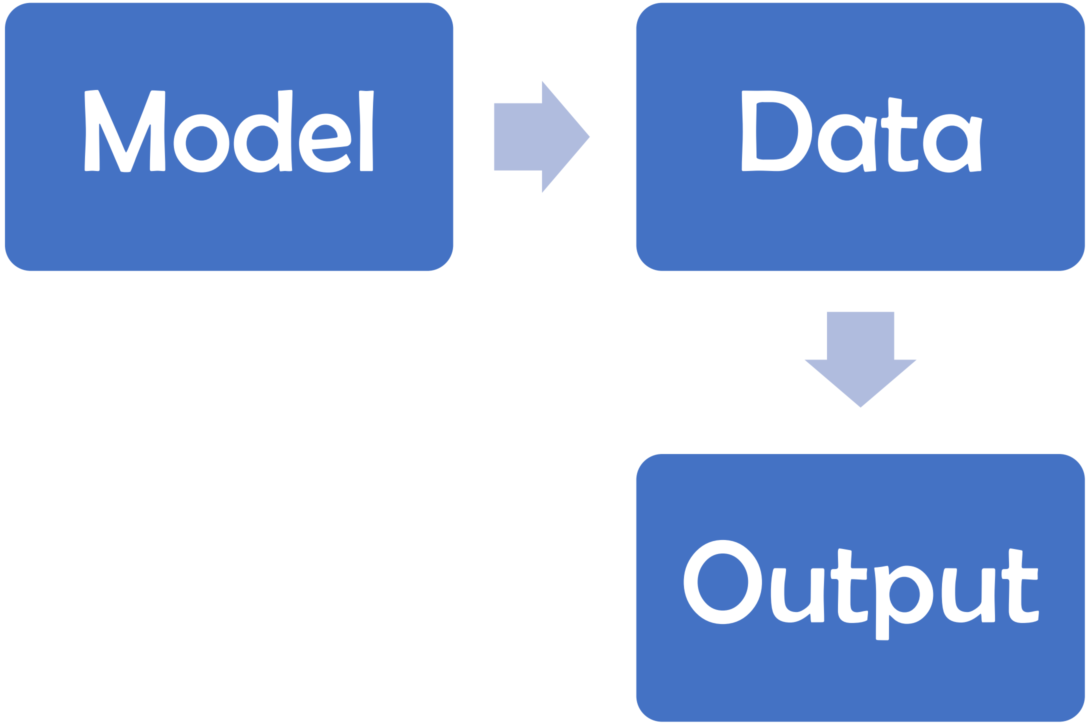
Adaptive data analysis

Classical setting

Adaptive setting

Adaptive data analysis

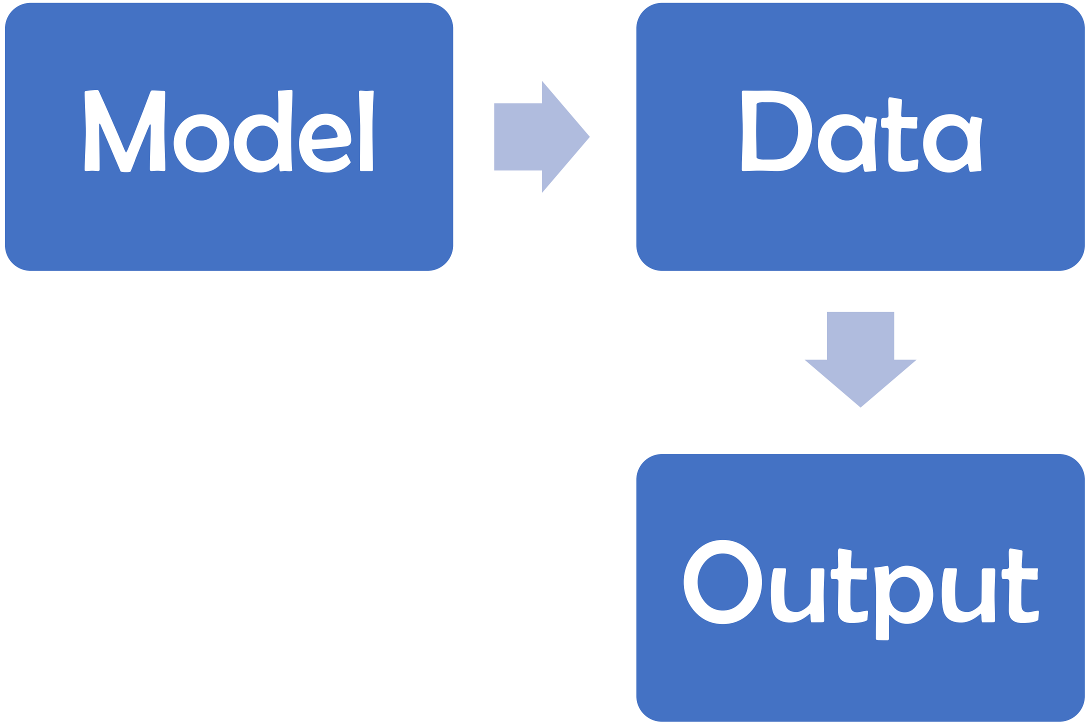
Classical setting



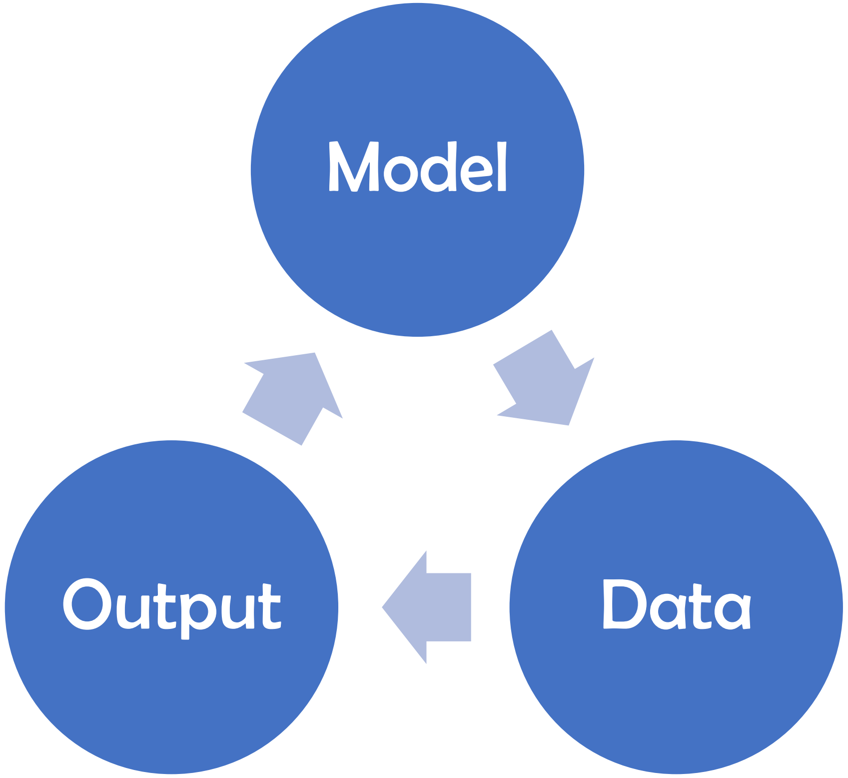
Adaptive setting

Adaptive data analysis

Classical setting

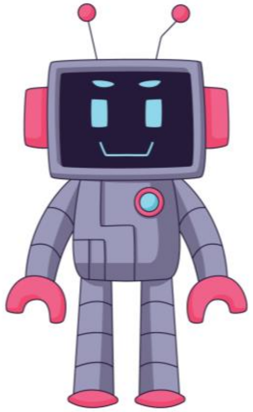


Adaptive setting



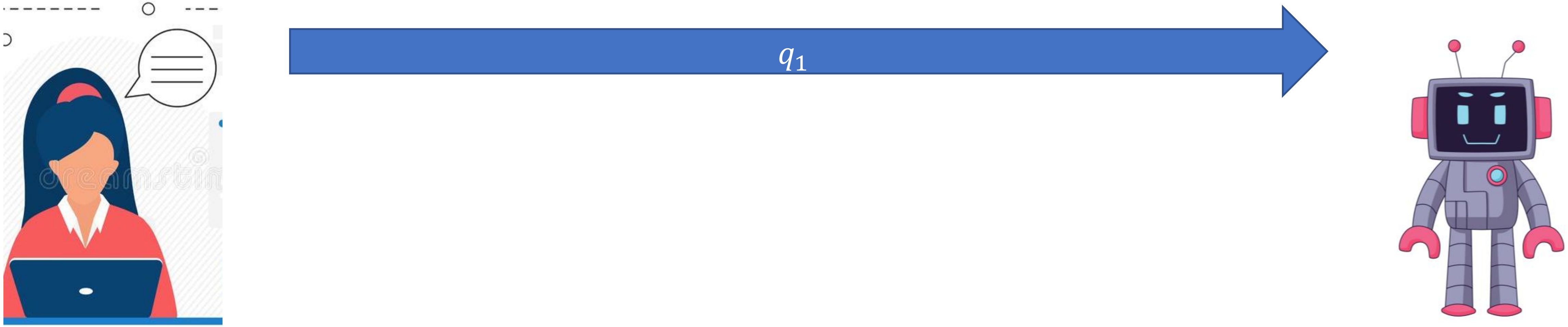
Adaptive data analysis – Formal(ish)

A



M

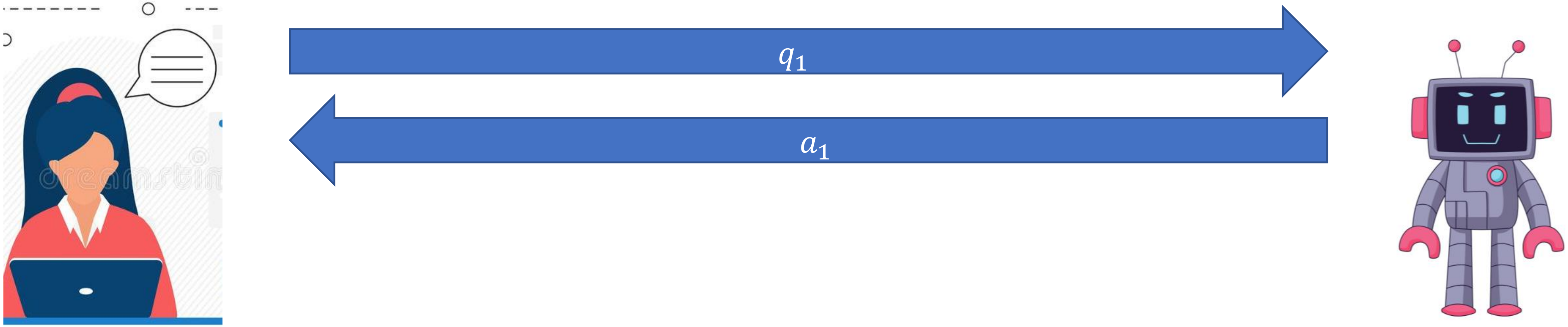
Adaptive data analysis – Formal(ish)



A

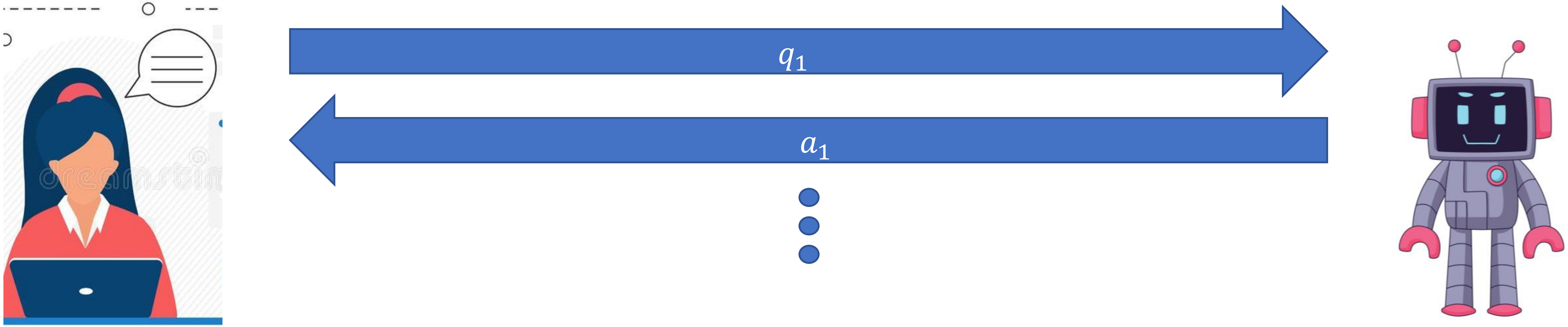
M

Adaptive data analysis – Formal(ish)



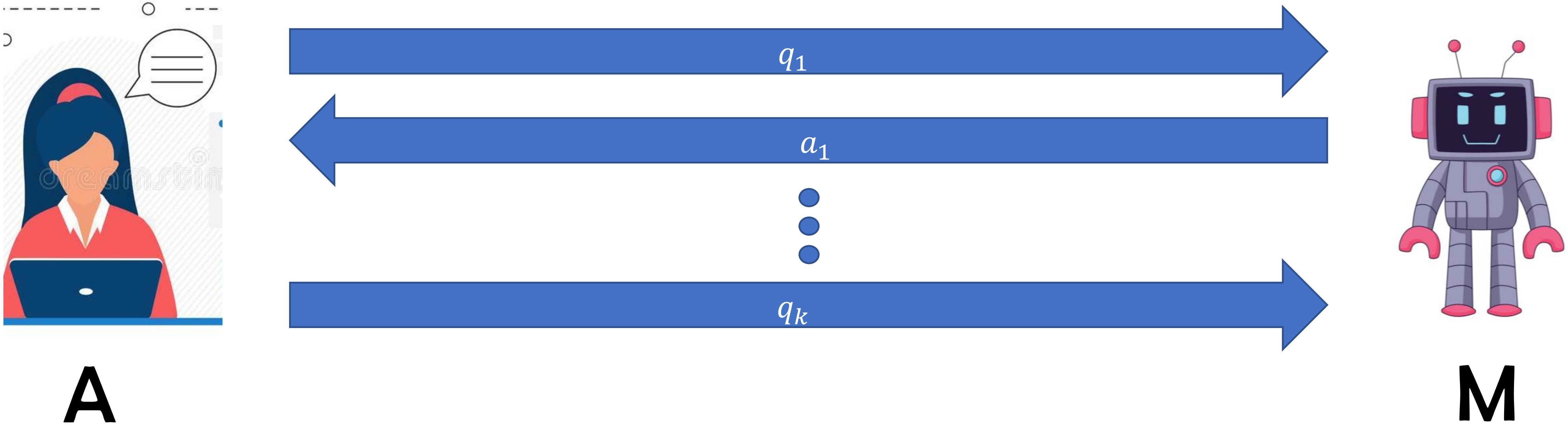
A M

Adaptive data analysis – Formal(ish)

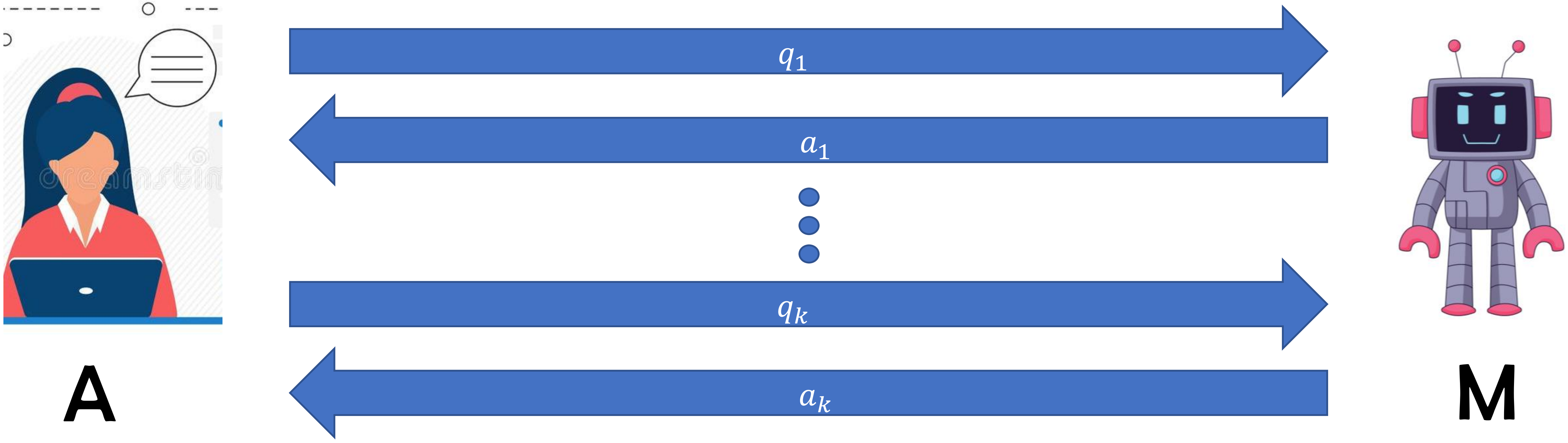


A M

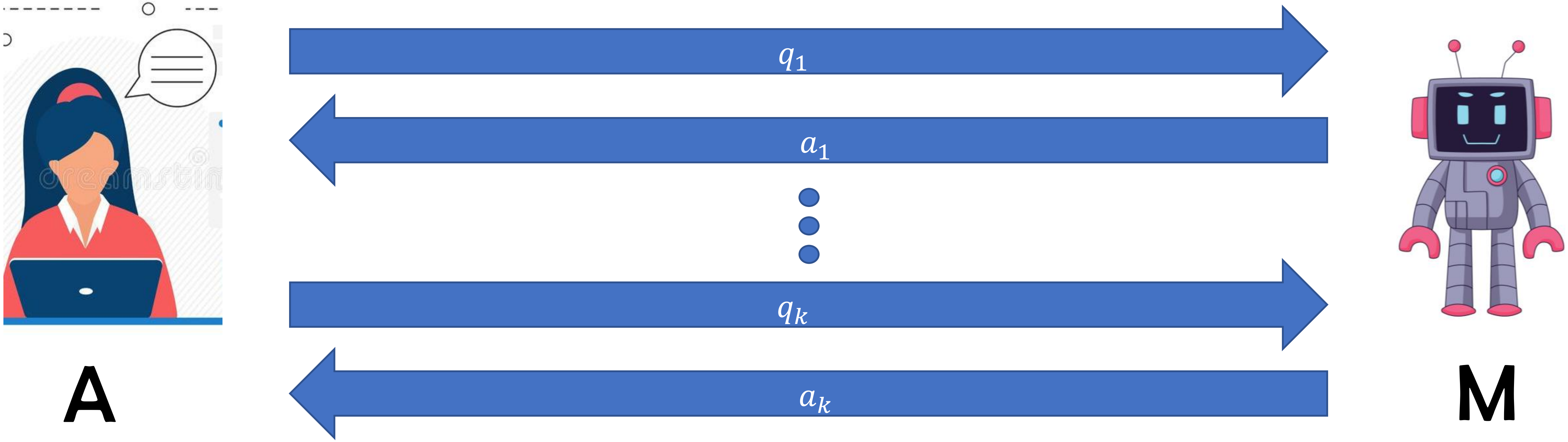
Adaptive data analysis – Formal(ish)



Adaptive data analysis – Formal(ish)



Adaptive data analysis – Formal(ish)



Goal:

- M tries to return accurate answers to the queries (with respect to the population).
- A tries to disrupt and find queries on which M will fail.

What is known

What is known

- Statistical validity can be preserved using

What is known

- Statistical validity can be preserved using
  - Differential privacy (Bassily et al., 2015)

What is known

- Statistical validity can be preserved using
  - Differential privacy (Bassily et al., 2015)
  - Transcription compression (Dwork et al. 2015a)

What is known

- Statistical validity can be preserved using
  - Differential privacy (Bassily et al., 2015)
  - Transcription compression (Dwork et al. 2015a)
  - Max-information (Dwork et al. 2015b)

What is known

- Statistical validity can be preserved using
  - Differential privacy (Bassily et al., 2015)
  - Transcription compression (Dwork et al. 2015a)
  - Max-information (Dwork et al. 2015b)
  - Typical stability (Bassily & Freund, 2016)

What is known

- Statistical validity can be preserved using
  - Differential privacy (Bassily et al., 2015)
  - Transcription compression (Dwork et al. 2015a)
  - Max-information (Dwork et al. 2015b)
  - Typical stability (Bassily & Freund, 2016)
  - Universal conditional mutual information (Steinke & Zakynthinou 2020)

What is known

- Statistical validity can be preserved using
  - Differential privacy (Bassily et al., 2015)
  - Transcription compression (Dwork et al. 2015a)
  - Max-information (Dwork et al. 2015b)
  - Typical stability (Bassily & Freund, 2016)
  - Universal conditional mutual information (Steinke & Zakynthinou 2020)
- All of which are heavily based on the i.i.d assumption.

Learning with correlated observations

Learning with correlated observations

- Classical statistics (Pearson, 1895; Terence 1990)

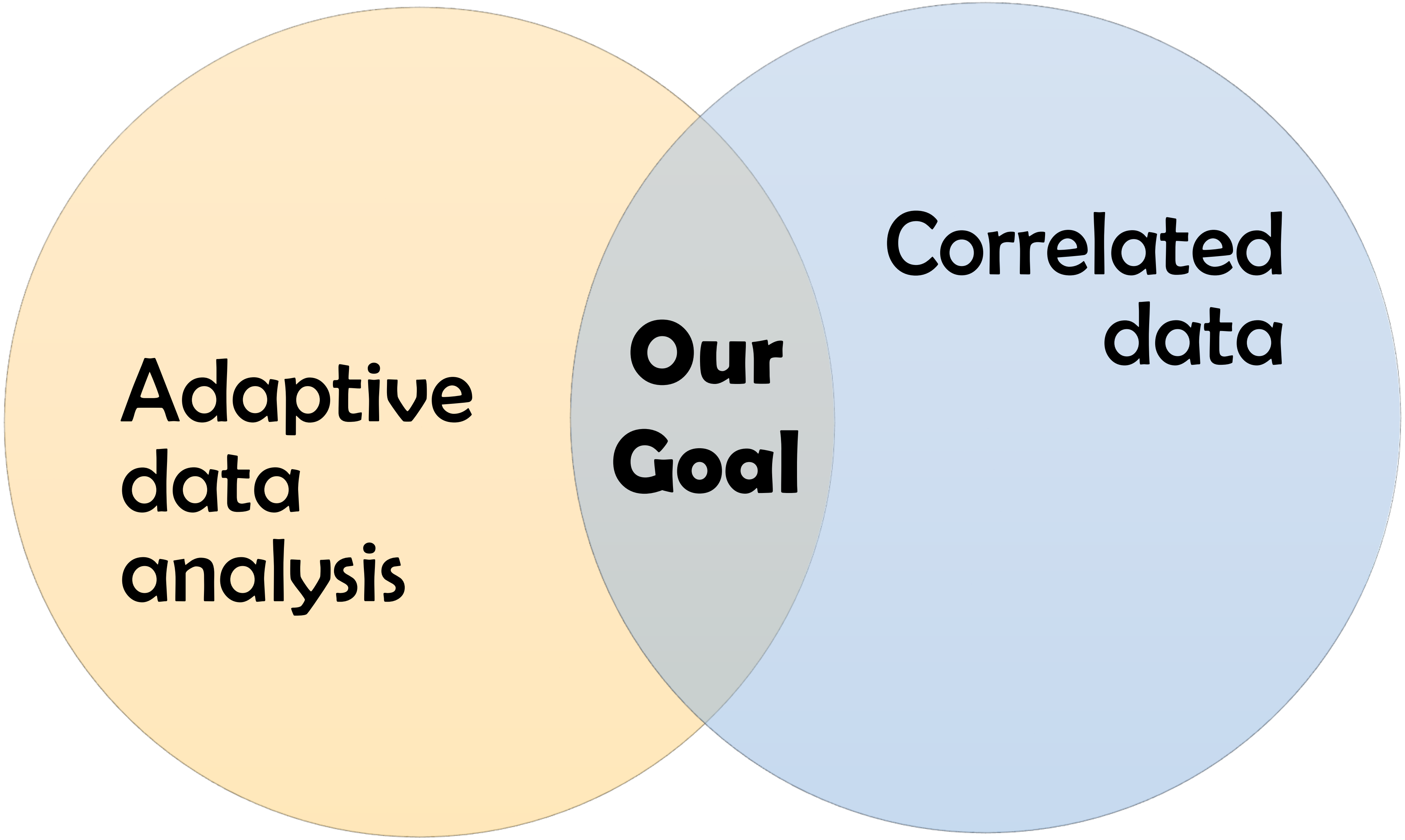
Learning with correlated observations

- Classical statistics (Pearson, 1895; Terence 1990)
- Mixing Markov chains (Marton, 1996; Kontorovich and Raginsky, 2017)

Learning with correlated observations

- Classical statistics (Pearson, 1895; Terence 1990)
- Mixing Markov chains (Marton, 1996; Kontorovich and Raginsky, 2017)
- Dobrushin condition (Dagan et al. 2019)

Our
Goal


Adaptive data analysis
Our Goal
Correlated data

We do this in two ways

We do this in two ways

Differential Privacy

We do this in two ways

Differential Privacy

Gibbs dependency

$$\psi(\mu) := \sup_{x \in \mathcal{X}^n} \mathbb{E}_i \left\| \mu_i(\cdot) - \mu_i\left(\cdot \mid x^{-i} \right) \right\|_{TV}.$$

μ has ψ-Gibbs dependency if $\psi(\mu) \le \psi$.

We do this in two ways

Differential Privacy

Gibbs dependency

$$\psi(\mu) := \sup_{x \in \mathcal{X}^n} \mathbb{E}_i \left\| \mu_i(\cdot) - \mu_i\left(\cdot \mid x^{-i} \right) \right\|_{TV}.$$

μ has ψ-Gibbs dependency if $\psi(\mu) \leq \psi$.

Theorem

If $\psi(\mu)$ is bounded, differentially private mechanisms exhibit generalization in the adaptive setting

We do this in two ways

Differential Privacy

Gibbs dependency

$$\psi(\mu) := \sup_{x \in \mathcal{X}^n} \mathbb{E}_i \left\| \mu_i(\cdot) - \mu_i\left(\cdot \mid x^{-i} \right) \right\|_{TV}.$$

μ has ψ-Gibbs dependency if $\psi(\mu) \leq \psi$.

Theorem

If $\psi(\mu)$ is bounded, differentially private mechanisms exhibit generalization in the adaptive setting

This is tight

We do this in two ways

Differential Privacy

Gibbs dependency

$$\psi(\mu) \coloneqq \sup_{x \in \mathcal{X}^n} \mathbb{E}_i \left\| \mu_i(\cdot) - \mu_i\left(\cdot \mid x^{-i} \right) \right\|_{TV}.$$

μ has ψ-Gibbs dependency if $\psi(\mu) \leq \psi$.

Theorem

If $\psi(\mu)$ is bounded, differentially private mechanisms exhibit generalization in the adaptive setting

This is tight

And can be applied to Markov chains!

We do this in two ways

Differential Privacy

Gibbs dependency

$$\psi(\mu) := \sup_{x \in \mathcal{X}^n} \mathbb{E}_i \left\| \mu_i(\cdot) - \mu_i\left(\cdot \mid x^{-i}\right) \right\|_{TV}.$$

μ has ψ-Gibbs dependency if $\psi(\mu) \le \psi$.

Theorem

If $\psi(\mu)$ is bounded, differentially private mechanisms exhibit generalization in the adaptive setting

This is tight

And can be applied to Markov chains!

Transcript Compression

We do this in two ways

Differential Privacy

Gibbs dependency

$$\psi(\mu) := \sup_{x \in \mathcal{X}^n} \mathbb{E}_i \left\| \mu_i(\cdot) - \mu_i(\cdot \mid x^{-i}) \right\|_{TV}.$$

μ has ψ-Gibbs dependency if $\psi(\mu) \leq \psi$.

Theorem

If $\psi(\mu)$ is bounded, differentially private mechanisms exhibit generalization in the adaptive setting

This is tight

And can be applied to Markov chains!

Transcript Compression

Definition (Bassily and Freund (2016))

For a query q and distribution μ, we denote by $\gamma(q, \mu, \delta)$ the length of the confidence interval around the expectation of q, with confidence level $1 - \delta$

We do this in two ways

Differential Privacy

Gibbs dependency

$\psi(\mu) := \sup_{x \in \mathcal{X}^n} \mathbb{E}_i \left\| \mu_i(\cdot) - \mu_i(\cdot \mid x^{-i}) \right\|_{TV}.$

μ has ψ-Gibbs dependency if $\psi(\mu) \leq \psi$.

Theorem

If $\psi(\mu)$ is bounded, differentially private mechanisms exhibit generalization in the adaptive setting

This is tight

And can be applied to Markov chains!

Transcript Compression

Definition (Bassily and Freund (2016))

For a query q and distribution μ, we denote by $\gamma(q, \mu, \delta)$ the length of the confidence interval around the expectation of q, with confidence level $1 - \delta$

Theorem

We can efficiently and adaptively answer k queries with $\alpha + \gamma(q_i, \mu, \delta)$ accuracy w.h.p

Where do we go from here?

Where do we go from here?

- More tools for this regime

Where do we go from here?

- More tools for this regime
- More applications

Where do we go from here?

- More tools for this regime
- More applications
- More types correlations

Thank you for listening

See you in the poster session