Where data should go?

Arguments in favor of **private** data:
- Privacy and security
- Regulatory compliance
- Competitive advantage

Arguments in favor of **public** data:
- Improved decision-making
- Less barriers for entry
- Innovation, research
Where data should go?

Arguments in favor of private data:
- Privacy and security
- Regulatory compliance
- Competitive advantage

Arguments in favor of public data:
- Improved decision-making
- Less barriers for entry
- Innovation, research

**Synthetic** data serves as a middle ground!
By 2030 synthetic data will completely overshadow real data in AI models.
Synthetic power systems datasets

Texas A&M University Grid Datasets
(from 37 to 80k+ bus networks)

PyPSA-Eur: synthetic dataset of
Europe covering the full ENTSO-E area

Synthetic Data of the National Electricity Market (Australia)

Why these datasets may not satisfy our needs?

- "[...] data bears no relation to the actual grid [...] except that generation and load profiles are similar, based on public data"
- "This test case represents a synthetic (fictitious) transmission".
- "This case is synthetic and does not model the actual grid"
Synthetic power systems datasets

Texas A&M University Grid Datasets (from 37 to 80k+ bus networks)

PyPSA-Eur: synthetic dataset of Europe covering the full ENTSO-E area

Synthetic Data of the National Electricity Market (Australia)

Why these datasets may not satisfy our needs?

- “[…] data bears no relation to the actual grid […] except that generation and load profiles are similar, based on public data”
- “This test case represents a synthetic (fictitious) transmission”
- “This case is synthetic and does not model the actual grid”
Differential privacy & optimization for synthetic power systems data

Real-world dataset → Calibrated noise → Privacy-preserving dataset → Post-processing optimization → Privacy-preserving and consistent dataset
Formalizing differential privacy (DP)

- Wind power records $y, y', y'', ... \in [0, 1]$
- For given $\alpha > 0$, records $y$ and $y'$ are $\alpha$-adjacent if $\|y - y'\| \leq \alpha$
- The goal is to obfuscate differences in records up to $\alpha$
Formalizing differential privacy (DP)

- Wind power records $y, y', y'', \ldots \in [0, 1]$
- For given $\alpha > 0$, records $y$ and $y'$ are $\alpha$-adjacent if $\| y - y' \| \leq \alpha$
- The goal is to obfuscate differences in records up to $\alpha$
Formalizing differential privacy (DP)

- Wind power records \( y, y', y'', \ldots \in [0, 1] \)
- For given \( \alpha > 0 \), records \( y \) and \( y' \) are \( \alpha \)-adjacent if \( ||y - y'|| \leq \alpha \)
- The goal is to obfuscate differences in records up to \( \alpha \)

- Let \( \zeta \sim \text{Lap}(\alpha/\varepsilon) \) be a zero-mean random Laplacian noise
- For some small parameter \( \varepsilon > 0 \), the release is \( \varepsilon \)-DP if

\[
\frac{\Pr[y + \zeta \in \hat{y}]}{\Pr[y' + \zeta \in \hat{y}]} \leq \exp(\varepsilon)
\]
Formalizing differential privacy (DP)

- Wind power records \( y, y', y'', \ldots \in [0, 1] \)
- For given \( \alpha > 0 \), records \( y \) and \( y' \) are \( \alpha \)-adjacent if \( \| y - y' \| \leq \alpha \)
- The goal is to obfuscate differences in records up to \( \alpha \)
- Let \( \zeta \sim \text{Lap}(\alpha/\varepsilon) \) be a zero-mean random Laplacian noise
- For some small parameter \( \varepsilon > 0 \), the release is \( \varepsilon \)-DP if

\[
\frac{\Pr[ y + \zeta \in \hat{y} ]}{\Pr[ y' + \zeta \in \hat{y} ]} \leq \exp(\varepsilon)
\]
Formalizing differential privacy (DP)

Wind power records $y, y', y'', \ldots \in [0, 1]$

For given $\alpha > 0$, records $y$ and $y'$ are $\alpha$-adjacent if $\|y - y'\| \leq \alpha$

The goal is to obfuscate differences in records up to $\alpha$

Let $\zeta \sim \text{Lap}(\alpha/\varepsilon)$ be a zero-mean random Laplacian noise

For some small parameter $\varepsilon > 0$, the release is $\varepsilon$-DP if

$$\frac{\Pr[y + \zeta \in \hat{y}]}{\Pr[y' + \zeta \in \hat{y}]} \leq \exp(\varepsilon)$$
Formalizing differential privacy (DP)

- Wind power records $y, y', y'', \ldots \in [0, 1]$
- For given $\alpha > 0$, records $y$ and $y'$ are $\alpha$-adjacent if $||y - y'|| \leq \alpha$
- The goal is to obfuscate differences in records up to $\alpha$

Strong theoretical properties
- Rigorous, quantifiable privacy guarantees
- Immunity to post-processing! Arbitrary transformations of noisy data preserve privacy
Wind power obfuscation (WPO) algorithm (Part I)

real dataset: \( \mathcal{D} = \{ (y_1, x_1), \ldots, (y_n, x_n) \} \)

synthetic dataset: \( \tilde{\mathcal{D}} = \{ (\tilde{y}_1, x_1), \ldots, (\tilde{y}_n, x_n) \} \)

- Regression on synthetic data \( \tilde{y} \) must match the regression on real data \( y \)
- We use regression loss and weights as a measure of accuracy
- Private estimation of regression parameters:

\[
\begin{align*}
\text{loss: } & \quad \ell = \ell(y) + \text{Lap}\left( \frac{\delta_\ell}{\epsilon} \right), \\
\text{weights: } & \quad \tilde{\beta} = \beta(y) + \text{Lap}\left( \frac{\delta_\beta}{\epsilon} \right)
\end{align*}
\]

where \( \delta(\cdot) \) is the sensitivity of \( \cdot \) to data \( \alpha \)-adjacent datasets

- Lemma (global sensitivity bounds):

\[
\delta_\ell \leq \max_{i=1,\ldots,n} \left\| (X(X^T X + \lambda I)^{-1} X^T - I)(e_i \circ \alpha) \right\| \quad \delta_\beta \leq \left\| (X^T X + \lambda I)^{-1} X^T \right\|_1 \alpha
\]
Wind power obfuscation (WPO) algorithm (Part I)

real dataset: \( \mathcal{D} = \{(y_1, x_1), \ldots, (y_n, x_n)\} \)

synthetic dataset: \( \tilde{\mathcal{D}} = \{ (\tilde{y}_1, x_1), \ldots, (\tilde{y}_n, x_n) \} \)

- Regression on synthetic data \( \tilde{y} \) must match the regression on real data \( y \)
- We use regression loss and weights as a measure of accuracy
- Private estimation of regression parameters:

  \[ \text{loss : } \bar{\ell} = \ell(y) + \text{Lap}\left( \frac{\delta_{\ell}}{\varepsilon} \right), \quad \text{weights : } \bar{\beta} = \beta(y) + \text{Lap}\left( \frac{\delta_{\beta}}{\varepsilon} \right) \]

  where \( \delta_{(\cdot)} \) is the sensitivity of \( (\cdot) \) to data \( \alpha \)-adjacent datasets

**Lemma (global sensitivity bounds):**

\[ \delta_{\ell} \leq \max_{i=1,\ldots,n} \| (X(X^TX + \lambda I)^{-1}X^T - I)(e_i \circ \alpha) \| \]

\[ \delta_{\beta} \leq \left\| (X^TX + \lambda I)^{-1}X^T \right\|_1 \alpha \]
Wind power obfuscation (WPO) algorithm (Part I)

real dataset: \( \mathcal{D} = \{(y_1, x_1), \ldots, (y_n, x_n)\} \)

synthetic dataset: \( \tilde{\mathcal{D}} = \{ (\tilde{y}_1, x_1), \ldots, (\tilde{y}_n, x_n) \} \)

- Regression on synthetic data \( \tilde{y} \) must match the regression on real data \( y \)
- We use regression loss and weights as a measure of accuracy
- Private estimation of regression parameters:

\[
\text{loss : } \ell = \ell(y) + \text{Lap} \left( \frac{\delta_\ell}{\varepsilon} \right), \quad \text{weights : } \beta = \beta(y) + \text{Lap} \left( \frac{\delta_\beta}{\varepsilon} \right)
\]

where \( \delta(\cdot) \) is the sensitivity of \( (\cdot) \) to data \( \alpha \)-adjacent datasets

- Lemma (global sensitivity bounds):

\[
\delta_\ell \leq \max_{i=1,\ldots,n} \left\| (X(X^T X + \lambda I)^{-1} X^T - I)(e_i \circ \alpha) \right\| \quad \delta_\beta \leq \left\| (X^T X + \lambda I)^{-1} X^T \right\|_1 \alpha
\]

\[
\text{min}_{\beta} \| X\beta - y \| + \lambda \| \beta \|
\]
Wind power obfuscation (WPO) algorithm (Part II)

**Step 1** Synthetic wind power measurements:

\[ \tilde{y}^0 = y + \text{Lap} \left( \frac{\alpha}{\varepsilon_1} \right) \]

**Step 2** Private regression parameters estimation:

\[ \overline{\ell} = \ell(y) + \text{Lap} \left( \frac{\delta_{\ell}}{2\varepsilon_2} \right) \quad \overline{\beta} = \beta(y) + \text{Lap} \left( \frac{\delta_{\beta}}{2\varepsilon_2} \right) \]

**Step 3** Synthetic dataset post-processing:

\[ \bar{y} \in \arg\min_{\tilde{y}} \quad \underbrace{\| \overline{\ell} - \ell(\tilde{y}) \|}_{\text{loss accuracy}} + \gamma_{\beta} \underbrace{\| \overline{\beta} - \beta(\tilde{y}) \|}_{\text{weight accuracy}} + \gamma_y \underbrace{\| \tilde{y}^0 - \tilde{y} \|}_{\text{regularization}} \]

s.t. \( 0 \leq \bar{y} \leq 1 \)

\[ \beta(\bar{y}), \ell(\bar{y}) \in \arg\min_{\beta} \quad \underbrace{\| X\beta - \tilde{y} \|}_{\ell} + \lambda \| \beta \| \]

**Theorem:** \( \varepsilon_1 = \varepsilon/2 \) and \( \varepsilon_2 = \varepsilon/4 \) renders WPO \( \varepsilon-\text{DP} \) for \( \alpha-\text{adjacent wind power datasets.} \)
Wind power obfuscation (WPO) algorithm (Part II)

**Step 1** Synthetic wind power measurements:
\[ \hat{y}^0 = y + \text{Lap}(\alpha / \varepsilon_1) \]

**Step 2** Private regression parameters estimation:
\[ \bar{\ell} = \ell(y) + \text{Lap}(\delta_{\ell} / \varepsilon_2) \quad \bar{\beta} = \beta(y) + \text{Lap}(\delta_{\beta} / \varepsilon_2) \]

**Step 3** Synthetic dataset post-processing:
\[ \hat{y} \in \text{argmin}_{\tilde{y}} \quad \left( ||\bar{\ell} - \ell(\tilde{y})|| + \gamma_{\beta} ||\bar{\beta} - \beta(\tilde{y})|| + \gamma_y ||\hat{y}^0 - \tilde{y}|| \right) \]

\[ \text{s.t.} \quad 0 \leq \hat{y} \leq 1 \]
\[ \beta(\hat{y}), \ell(\hat{y}) \in \text{argmin}_{\beta} \left( \|X\beta - \hat{y}\| + \lambda \|\beta\| \right) \]

**Theorem:** \( \varepsilon_1 = \varepsilon / 2 \) and \( \varepsilon_2 = \varepsilon / 4 \) renders WPO \( \varepsilon \)-DP for \( \alpha \)-adjacent wind power datasets.
WPO algorithm: Application to Alstom Eco 80 wind turbine

Laplace Mechanism

Accuracy of the WPO Algorithm remains high with a growing privacy requirement $\alpha$
Differentially private release of network parameters

**Optimal Power Flow (OPF) problem**

\[
C(\bar{f}) = \min_{p \in \mathcal{P}} \quad c^T p \\
\text{s.t.} \quad 1^T (p - d) = 0 \\
|F(p - d)| \leq \bar{f}
\]

*dispatch costs*

*power balance*

*power flow limit*

How to release vector of transmission capacities \( \bar{f} \) privately?

**Laplace mechanism:**

\[
\bar{\varphi}^0 = \bar{f} + \text{Lap}(\alpha/\varepsilon)
\]

Almost never feasible

**Laplace + Bilevel optimization:**

\[
\min_{\hat{\varphi}} \quad \|\bar{\varphi}^0 - \hat{\varphi}\| \\
\text{s.t.} \quad |C(\hat{\varphi}) - C^*| \leq \beta C^*
\]

Feasible and cost-consistent with respect to a single OPF model

**Laplace & Exponential mechanisms + Bilevel optimization:**

- LM for obfuscation
- EM for worst-case OPF models
- Bilevel opt. on worst-case models

Feasible and cost-consistent with respect to a population of OPF models
Optimal Power Flow (OPF) problem

\[
C(\mathbf{f}) = \min_{\mathbf{p} \in \mathcal{P}} c^T \mathbf{p} \\
\text{s.t.} \quad 1^T (\mathbf{p} - d) = 0 \\
|F(\mathbf{p} - d)| \leq \mathbf{f}
\]

dispatch costs

power balance

power flow limit

How to release vector of transmission capacities \( \mathbf{f} \) privately?

Laplace mechanism:

\[
\varphi^0 = \mathbf{f} + \text{Lap}(\alpha/\varepsilon)
\]

Almost never feasible

Laplace + Bilevel optimization:

\[
\min_{\hat{\varphi}} \| \varphi^0 - \hat{\varphi} \| \\
\text{s.t.} \quad |C(\hat{\varphi}) - C^*| \leq \beta C^*
\]

Feasible and cost-consistent with respect to a single OPF model

Laplace & Exponential mechanisms + Bilevel optimization:

- LM for obfuscation
- EM for worst-case OPF models
- Bilevel opt. on worst-case models

Feasible and cost-consistent with respect to a population of OPF models
Differentially private release of network parameters

**Optimal Power Flow (OPF) problem**

\[
C(\overline{f}) = \min_{p \in \mathcal{P}} \quad c^T p \\
\text{s.t.} \quad 1^T (p - d) = 0 \\
\quad |F(p - d)| \leq \overline{f}
\]

dispatch costs

power balance

power flow limit

How to release vector of transmission capacities \(\overline{f}\) privately?

**Laplace mechanism:**

\[
\varphi^0 = \overline{f} + \text{Lap}(\alpha/\varepsilon)
\]

Almost never feasible

**Laplace + Bilevel optimization:**

\[
\min_{\hat{\varphi}} \quad \| \varphi^0 - \hat{\varphi} \| \\
\text{s.t.} \quad |C(\hat{\varphi}) - C_*| \leq \beta C_*
\]

Feasible and cost-consistent with respect to a **single** OPF model

**Laplace & Exponential mechanisms + Bilevel optimization:**

- LM for obfuscation
- EM for worst-case OPF models
- Bilevel opt. on worst-case models

Feasible and cost-consistent with respect to a **population** of OPF models
Differentially private release of network parameters

Optimal Power Flow (OPF) problem

\[ C(\bar{f}) = \min_{p \in \mathcal{P}} \quad c^T p \]
\[ \text{subject to} \quad 1^T (p - d) = 0 \]
\[ |F(p - d)| \leq \bar{f} \]

\( \text{dispatch costs} \)
\( \text{power balance} \)
\( \text{power flow limit} \)

How to release vector of transmission capacities \( \bar{f} \) privately?

Laplace mechanism:

\[ \bar{\varphi}^0 = \bar{f} + \text{Lap}(\alpha/\varepsilon) \]

Almost never feasible

Laplace + Bilevel optimization:

\[ \min_{\hat{\varphi}} \quad \| \bar{\varphi}^0 - \hat{\varphi} \| \]
\[ \text{subject to} \quad |C(\hat{\varphi}) - C^*| \leq \beta C^* \]

Feasible and cost-consistent with respect to a single OPF model

Laplace \& Exponential mechanisms + Bilevel optimization:

- LM for obfuscation
- EM for worst-case OPF models
- Bilevel opt. on worst-case models

Embedded OPF

Feasible and cost-consistent with respect to a population of OPF models
**Differentially private transmission capacity obfuscation (TCO) Algorithm**

**Step 1** Initialize synthetic data using LM:

\[
\bar{\varphi}^0 = \bar{f} + \text{Lap}(\alpha/\varepsilon_1)
\]

**Step 2** Find the worst-case OPF model using EM:

\[
\Delta C_i = \left\| C_i(\bar{f}) - C_i^R(\bar{\varphi}^0) \right\|_1 + \text{Lap}(\bar{c}\alpha/\varepsilon_2), \forall i = 1, \ldots, m
\]

return index \(k\) of the worst-case model

**Step 3** Compute the worst-case cost using LM:

\[
\bar{c} = C_k(\bar{f}) + \text{Lap}(\bar{c}\alpha/\varepsilon_2)
\]

**Step 4** Post-processing bilevel optimization:

\[
\bar{\varphi}^* \in \arg\min_{\varphi} \left\| \bar{c} - C_k(\varphi) \right\| + \left\| \bar{\varphi} - \bar{\varphi}^0 \right\|
\]

**Theorem:** \(\varepsilon_1 = \varepsilon/2\) and \(\varepsilon_2 = \varepsilon/(4T)\) achieve \(\varepsilon\)—differential privacy
Differentially private transmission capacity obfuscation (TCO) Algorithm

**Step 1** Initialize synthetic data using LM:

\[ \varphi^0 = \overline{f} + \text{Lap}(\alpha/\varepsilon_1) \]

**Step 2** Find the worst-case OPF model using EM:

\[ \Delta C_i = \left\| C_i(\overline{f}) - C^R_i(\varphi^0) \right\|_1 + \text{Lap}(\overline{c}\alpha/\varepsilon_2), \forall i = 1, \ldots, m \]

return index \( k \) of the worst-case model

**Step 3** Compute the worst-case cost using LM:

\[ \overline{C} = C_k(\overline{f}) + \text{Lap}(\overline{c}\alpha/\varepsilon_2) \]

**Step 4** Post-processing bilevel optimization:

\[ \varphi^* \in \text{argmin} \| \overline{C} - C_k(\varphi) \| + \| \varphi - \varphi^0 \| \]

**Theorem:** \( \varepsilon_1 = \varepsilon/2 \) and \( \varepsilon_2 = \varepsilon/(4T) \) achieve \( \varepsilon \)-differential privacy
Differentially private transmission capacity obfuscation (TCO) Algorithm

**Step 1** Initialize synthetic data using LM:

\[ \varphi^0 = \bar{f} + \text{Lap}(\alpha/\varepsilon_1) \]

**Step 2** Find the worst-case OPF model using EM:

\[ \Delta C_i = \left\| C_i(\bar{f}) - C_i^R(\varphi^0) \right\|_1 + \text{Lap}(\bar{\alpha}/\varepsilon_2), \forall i = 1, \ldots, m \]

return index \( k \) of the worst-case model

**Step 3** Compute the worst-case cost using LM:

\[ \bar{C} = C_k(\bar{f}) + \text{Lap}(\bar{\alpha}/\varepsilon_2) \]

**Step 4** Post-processing bilevel optimization:

\[ \varphi^* \in \text{argmin}_{\varphi} \left\| \bar{C} - C_k(\varphi) \right\| + \left\| \varphi - \varphi^0 \right\| \]

**Theorem:** \( \varepsilon_1 = \varepsilon/2 \) and \( \varepsilon_2 = \varepsilon/(4T) \) achieve \( \varepsilon \)-differential privacy
Differentially private transmission capacity obfuscation (TCO) Algorithm

**Step 1** Initialize synthetic data using LM:

$$\varphi^0 = \bar{f} + \text{Lap}(\alpha/\varepsilon)$$

**Step 2** Find the worst-case OPF model using EM:

$$\Delta C_i = \| C_i(\bar{f}) - C_i^R(\varphi^0)\|_1 + \text{Lap}(\bar{c}\alpha/\varepsilon_2), \forall i = 1, \ldots, m$$

return index $k$ of the worst-case model

**Step 3** Compute the worst-case cost using LM:

$$\bar{C} = C_k(\bar{f}) + \text{Lap}(\bar{c}\alpha/\varepsilon_2)$$

**Step 4** Post-processing bilevel optimization:

$$\varphi^* \in \text{argmin} \| \bar{C} - C_k(\bar{\varphi})\| + \|\bar{\varphi} - \varphi^0\|$$

Theorem: $\varepsilon_1 = \varepsilon/2$ and $\varepsilon_2 = \varepsilon/(4T)$ achieve $\varepsilon$—differential privacy
Differentially private transmission capacity obfuscation (TCO) Algorithm

**Step 1** Initialize synthetic data using LM:

\[ \varphi^0 = f + \text{Lap}(\alpha/\varepsilon_1) \]

**Step 2** Find the worst-case OPF model using EM:

\[ \Delta C_i = \left\| C_i(f) - C_i^R(\varphi^{t-1}) \right\|_1 + \text{Lap}(\bar{c}\alpha/\varepsilon_2) \quad \forall i = 1, \ldots, m \]

return index \( k^t \) of the worst-case model

**Step 3** Compute the worst-case cost using LM:

\[ \bar{C}_t = C_k^t(f) + \text{Lap}(\bar{c}\alpha/\varepsilon_2) \]

**Step 4** Post-processing bilevel optimization:

\[ \varphi^t \in \text{argmin}_{\varphi} \sum_{\tau=1}^t \left\| \bar{C}_\tau - C_k^\tau(\varphi) \right\| + \left\| \varphi - \varphi^{t-1} \right\| \]

**Theorem:** \( \varepsilon_1 = \varepsilon/2 \) and \( \varepsilon_2 = \varepsilon/(4T) \) achieve \( \varepsilon \)-differential privacy
**Differentially private transmission capacity obfuscation (TCO) Algorithm**

**Step 1** Initialize synthetic data using LM:

\[
\varphi^0 = \overline{f} + \text{Lap}(\alpha/\varepsilon_1)
\]

**Step 2** Find the worst-case OPF model using EM:

\[
\Delta C_i = \left\| C_i(\overline{f}) - C^R_i(\varphi^{t-1}) \right\|_1 + \text{Lap}(\overline{c}\alpha/\varepsilon_2), \forall i = 1, \ldots, m
\]

return index \(k^t\) of the worst-case model

**Step 3** Compute the worst-case cost using LM:

\[
\overline{C}_t = C_{k^t}(\overline{f}) + \text{Lap}(\overline{c}\alpha/\varepsilon_2)
\]

**Step 4** Post-processing bilevel optimization:

\[
\varphi^t \in \arg\min \sum_{\tau=1}^t \left\| \overline{C}_\tau - C_{k^\tau}(\overline{\varphi}) \right\| + \left\| \overline{\varphi} - \overline{\varphi}^{t-1} \right\|
\]

**Theorem:** \(\varepsilon_1 = \varepsilon/2\) and \(\varepsilon_2 = \varepsilon/(4T)\) achieve \(\varepsilon\)–differential privacy
IEEE 73-RTS benchmark: Laplace versus TCO Algorithm

Laplace mechanism

- infeas: 100.0%
- suboptimality: 14.2%

TCO Algorithm

- iteration: 1
- infeas: 98.0%
- suboptimality: 11.4%
IEEE 73-RTS benchmark: TCO feasibility and sup-optimality

more privacy required

adjacency $\alpha = 5$MW

adjacency $\alpha = 15$MW

adjacency $\alpha = 30$MW

infeasibility [%]

iteration limit $T$
more privacy required

IEEE 73-RTS benchmark: TCO feasibility and sup-optimality

adjacency $\alpha = 5$MW

adjacency $\alpha = 15$MW

adjacency $\alpha = 30$MW

infeasibility [%]

iteration limit $T$

$\Delta C [%]$

iteration limit $T$

iteration limit $T$
IEEE 73-RTS benchmark: TCO robustness bias

The figure shows a box plot of the capacity of line #40 [MW] after 1, 5, and 10 iterations, with different values of $\alpha$: 5MW, 15MW, and 30MW, and the true capacity represented by a green line. The distribution of capacities varies across iterations and values of $\alpha$.
IEEE 73-RTS benchmark: TCO robustness bias

The diagram illustrates the capacity of line #40 [MW] after 1 iteration, after 5 iterations, and after 10 iterations, under different values of $\alpha$: 5MW, 15MW, and 30MW, compared to the true value (green line).

- After 1 iteration:
  - $\alpha=5MW$: Capacity distribution
  - $\alpha=15MW$: Capacity distribution
  - $\alpha=30MW$: Capacity distribution

- After 5 iterations:
  - $\alpha=5MW$: Capacity distribution
  - $\alpha=15MW$: Capacity distribution
  - $\alpha=30MW$: Capacity distribution

- After 10 iterations:
  - $\alpha=5MW$: Capacity distribution
  - $\alpha=15MW$: Capacity distribution
  - $\alpha=30MW$: Capacity distribution
IEEE 73-RTS benchmark: TCO robustness bias
Future of synthetic power system datasets

What we used to say about synthetic datasets:

- “[...] data bears no relation to the actual grid [...]”
- “This test case represents [...] fictitious transmission”
- “This case is synthetic and does not model the actual grid”

What we will say about synthetic datasets:

- “This synthetic dataset is produced based on the data from a real-world power grid”
- “It is not possible to infer the real data from this synthetic dataset”
- “Computational results on this data are consistent with the real data”
What does it mean for electricity market operators?

- New algorithms for **controllable** market transparency:
  - infrastructure data (grid topology, network parameters, generation, loads, etc.)
  - market participation data (bidding quantities, prices, etc.)

- No need for aggregation:
  - system cost/load $\rightarrow$ nodal cost/load
  - aggregated generation $\rightarrow$ highly granular generation records

- Rigorous privacy quantification $\rightarrow$ legal compliance (e.g., US Census Bureau)
Where data should go?

Our $\varepsilon$—differentially private algorithms provide a non-discrete answer to this question!
Thank you for your attention!

From this talk:
1. Dvorkin, V., Botterud A.
   **Differentially private algorithms for synthetic power system datasets**
   IEEE Control Systems Letters, 2023

Other references:
2. Dvorkin, V., Fioretto, F., Van Hentenryck, P., Kazempour, J. and Pinson, P.
   **Privacy-preserving convex optimization: When differential privacy meets stochastic programming**
   **Differentially private optimal power flow for distribution grids**
   IEEE Transactions on Power Systems, 2021
   🏆 Best 2019–2021 Paper Award
4. Dvorkin, V., Van Hentenryck, P., Kazempour, J. and Pinson P.
   **Differentially private distributed optimal power flow**
   2020 Conference on Decision and Control

Let’s stay in touch:

Twitter: DvorkinVladimir
LinkedIn: Vladimir-Dvorkin
Email: dvorkin@mit.edu