ESSCaSS 2022

Behind the Scenes: How Does One Become a (Machine Learning) Researcher, and What Does It Mean To Be One?"

Christoph Lampert
Publicly-Funded Research Institute
• PhD-granting graduate school
• no undergraduate studies (but internships)
• founded in 2009
• located close to Vienna, Austria

Focus on
• curiosity-driven basic research
• interdisciplinarity: Computer Science, Mathematics, Biology, Physics, Chemistry, Earth&Climate Sciences, Neuroscience

Fully English-speaking
Science is Everywhere
Scientists are in High Demand
Scientific Career Steps in Academia

Standardized career path world-wide:
• Step 1: Obtain a Bachelor’s and/or Master’s Degree
• Step 2: Obtain a Doctorate/PhD
• Step 3: Work as a “Postdoc” for a few years
• Step 4: Become an Assistant Professor
• Step 5: Become a Tenured Professor

What about science in industry?
• Leave the process anywhere after Step 2
Case Study: me

• 2000: Masters degree in Pure Mathematics
  University of Bonn, Germany

• 2001: Research stay
  Chalmers University, Gothenburg, Sweden

• 2003: PhD in Pure Mathematics
  University of Bonn, Germany

• 2004-2007: Postdoc in Computer Science
  DFKI Kaiserslautern, Germany

• 2006: Internship at Google
  Mountain View, California, USA

• 2007-2010: Postdoc in Computer Science
  MPI Tübingen, Germany

• 2010-2015: Assistant Professor
  ISTA, Vienna, Austria

• since 2015: Professor
  ISTA, Vienna, Austria
What's noteworthy?

- **2000:** Masters degree in Pure Mathematics
  University of Bonn, Germany

- **2003:** PhD in Pure Mathematics
  University of Bonn, Germany

- **2004-2007:** Postdoc in Computer Science
  DFKI Kaiserslautern, Germany

- **2006:** Internship at Google
  Mountain View, California, USA

- **2007-2010:** Postdoc in Computer Science
  MPI Tübingen, Germany

- **2010-2015:** Assistant Professor
  ISTA, Vienna, Austria

- **since 2015:** Professor
  ISTA, Vienna, Austria

I completed all steps.
What's noteworthy?

- 2000: Masters degree in Pure Mathematics
  University of Bonn, Germany
- 2003: PhD in Pure Mathematics
  University of Bonn, Germany
- 2004-2007: Postdoc in Computer Science
  DFKI Kaiserslautern, Germany
- 2006: Internship at Google
  Mountain View, California, USA
- 2007-2010: Postdoc in Computer Science
  MPI Tübingen, Germany
- 2010-2015: Assistant Professor
  ISTA, Vienna, Austria
- since 2015: Professor
  ISTA, Vienna, Austria

I completed all steps.
I switched fields after the PhD.
before:

Complex Analysis

Lemma 72. Wir untersuchen verschiedene Fälle für \( E_n \) analog zu Satz 73. Es gilt dazu

\[
\begin{align*}
2M^2 E_n &= E_n, \\
2M^2 \sigma B \wedge E_n(n, \cdot) &= (1 - |\tau|^2)\sigma B \wedge E_n, \\
2M^2 \sigma \bar{B} \wedge E_n(n, \cdot) &= \frac{\bar{\nu}}{|\bar{\nu}|} \sigma B \wedge E_n, \\
2M^2 \sigma \bar{B} \wedge E_n(n, \cdot) &= \frac{\bar{\tau}}{|\bar{\tau}|} \sigma B \wedge E_n.
\end{align*}
\]

Beweis. Der Beweis ist jeweils durch Einsetzung der Substitutionen \( M^2 \) und \( \sigma \).

Schließlich gelangen wir zum eigentlichen Ergebnis dieses Abschnitts.

Satz 73. Es sei \( A \) zulässig und von einer Form, so daß \( A' \) definiert und transformiert zulässig ist. Der Term \( E_n \) sei zeigte in

\[ E_n = B^1 + \delta B^1 \wedge B^2 + \omega B^2 \wedge B^3 + \delta B^3 \wedge B^4 + \delta B^4 \wedge B^5, \]

wobei \( B^1 \) bis \( B^5 \) jeweils identisch sind, nicht von \( \delta \) oder \( \omega \) abhängen und auch nicht die Differentiale \( \delta \) oder \( \omega \) enthalten. Dann ist

\[ A' = \frac{-1}{|\tau|^2} \left( (1 - |\tau|^2) \rho \wedge |\tau|^2 \rho \right) + (\rho \wedge |\tau|^2 \rho \wedge |\tau|^2 \rho \wedge |\tau|^2 \rho) + (\rho \wedge |\tau|^2 \rho \wedge |\tau|^2 \rho \wedge |\tau|^2 \rho). \]

Da sind \( P_1 \) bis \( P_3 \) Polynome in \( (1 - |\tau|^2)^2 \) zum Gauß \( -a - \delta \), normiert durch \( P_2(1) = 1 \).

Beweis. Ende der Substitutionen kombinieren, wissen wir bereits, wie \( \mathcal{A} \mathcal{L} \mathcal{A} \) aussieht. Für den Beweis verbleibt also nur von die Integration \( \mathcal{Z}^2 \) auszuführen. Die Rechnungen sind sehr ähnlich zu denen in \( \text{[LM90]} \), und die wir bereits an anderen Stellen verwiesen haben. Zunächst fokussieren wir alle Termen, bis nur noch Ausdrücke, die \( \sigma \) enthalten, unter dem Integrale an. Exemplarisch führen wir dies für die Terme mit \( E_n \) und \( \sigma \wedge E_n \) durch, für welche wir erhalten:

\[
\begin{align*}
\frac{2M}{e^2} \sigma^2 \wedge E_n &= \frac{a - 1}{e^2} \left( (1 - |\tau|^2)^2 \rho \wedge |\tau|^2 \rho \wedge |\tau|^2 \rho \wedge |\tau|^2 \rho \right), \\
\frac{2M}{e^2} \sigma \bar{B} \wedge E_n &= \frac{a - 1}{e^2} \left( (1 - |\tau|^2)^2 \rho \wedge |\tau|^2 \rho \wedge |\tau|^2 \rho \wedge |\tau|^2 \rho \right).
\end{align*}
\]

after:

Computer Vision

Figure 1: Quality of individual attribute predictions obtained on real data, mixed set classes, as measured by area under ROC curve (AUROC). Ambiguities with one class have constant values for all real classes, so that AUROC is undefined.

6. Conclusion

In this paper, we have introduced learning for distinct training and test data. We formulated the problem of learning an objective classification system for classes, for which no training images are available. We have proposed two methods for attribute-based classification that solve this problem by transforming it into a multi-class classification. The main idea is achieved in an intermediate representation that consists of high level semantic, per-class attributes, providing a fast and easy way to include human-like visual systems into the system.

Once trained, the system can detect any object category, for which a suitable characteristic by attributes is available, and it does not require a re-training step.

Additionally, we have introduced the "Attribute-cliques". We defined a set of attributes and a semantic similarity measure that can be used to group concepts in their cognitive relevance. We hope that this dataset will facilitate research and serve as a testbed for attribute-based classifiers.

Starting from this proposed system, various improvements and extensions are possible. Firstly, better designed per-attribute and multi-class classifiers could improve the overall performance of the system as it could be possible to choose the most suitable feature selection set based on the currently available data. Secondly, the introduction of a new attribute classifier could further improve the overall accuracy of the system. Finally, the proposed system could be extended to include human-like attributes, which could help to improve the overall performance of the system.
before:
Complex Analysis

Lemma 72. Wir untersuchen verschiedene Fälle für $E_{1t}$ analog zu Satz 73. Es gilt dann
- $2M_{t}E_{1t} = E_{1t}$
- $2M_{t}g_{t}E_{1t}(z_i) = (1 - |z_i|^2)g_{t}E_{1t}$
- $2M_{t}e_{t}E_{1t}(z_i) = \frac{\partial}{\partial r}g_{t}E_{1t}$
- $2M_{t}g_{t}E_{1t}(z_i) = \frac{\partial}{\partial r}g_{t}E_{1t}$
- $2M_{t}e_{t}E_{1t}(z_i) = \frac{\partial}{\partial r}e_{t}E_{1t}$

Beweis. Der Beweis ist jeweils direkt ausführen der Substitutionen $M_{t}$ und $R_{t}$.

Schließlich gelangen wir zum eigentlichen Ergebnis dieses Abschnitts

Satz 73. Es sei $A$ zulässig und von einer Form, so daß $A'$ definiert und transformiert zulässig ist. Der Term $E_{1t}$ sei definiert als
$$E_{1t} = E^{1} + \delta g_{t}E^{2} + \omega_{t}\delta g_{t}E^{3} + \delta e_{t}E^{4} + \delta e_{t}E^{5},$$

wobei $E^{1}$ bis $E^{5}$ jeweils invariant sind, nicht von $E$ oder $w_{t}$ abhängen und nicht die Differenzen $\delta g_{t}$ oder $\delta e_{t}$ enthalten. Dann ist
$$A' = \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{|A|}} e^{-\frac{1}{2}|A|^2} \left( A^{1} + \delta g_{t}E^{2} + \omega_{t}\delta g_{t}E^{3} + \delta e_{t}E^{4} + \delta e_{t}E^{5} \right).$$

Dabei sind $P_{t}$ bis $P_{t}$ Polynome in $\pi - 2\pi$ zum Grad $j - \alpha$, normalisiert durch $P_{t}(1) = 1$.

Beweis. Leiten wir die Substitutionsverfahren, wissen wir bereits, wie $M_{t}/\delta E_{1t}$ ausfällt. Für den Beweis verbleibt also nur die Integration $\delta F_{1t}$ auszuführen. Die Rechnungen sind sehr ähnlich zu denen in [Lam04], und die wir bereits an anderen Stellen verwiesen haben. Zunächst fokussieren wir alle Terme bis auf die Differenzen, die $\delta$ enthalten, und unter den Integralen stehen. Exemplarisch führen wir dann für die Terme $E_{1t}$ und $E_{1t}$ durch, für welche wir erhalten
$$2M_{t}E_{1t} = \frac{\partial}{\partial r}g_{t}E_{1t}(z_i)$$

also after:
Machine Learning

Maximum Margin Multi-Label Structured Prediction Supplemental Material

1 Generalization Properties of MILSP

We provide the proof of (Theorem 1 (Section 3.2) of the original manuscript).

Let $G_{1}(x) := \frac{1}{|y|} \sum_{x \neq y} g_{1}(x, y)$ be $\alpha$-piecewise linear. We assume $|y| \leq 1$ and $|y| \leq \alpha$ for all $x, y \in \mathbb{X}$, and $x \leq \alpha$ for all $y \in \mathbb{X}$. For any distribution, $\mathcal{Q}_{D_{t}}$, over weight vectors, that can approximate $\alpha$, is shown by $\mathcal{D}_{t}(\mathcal{Q}_{D_{t}})$ for the expected $\mathcal{D}_{t}$-risk for $\alpha$-dimensional data.

$$\mathcal{D}_{t}(\mathcal{Q}_{D_{t}}) = K_{D_{t}} - K_{D_{t}, \mathcal{Q}_{D_{t}}} - \mathbb{E}_{\mathcal{D}_{t}}[||\mathcal{Q}_{D_{t}}||_{2}]$$

Theorem 1. With probability at least $1 - \alpha$ over the sample $\mathcal{S}$ of the following inequality holds simultaneously for all weight vectors $w_{t}$:

$$L(\mathcal{Q}_{D_{t}}, \mathcal{S}) \leq \sum_{(x, y) \in \mathcal{S}} L(x, y, f, \mathcal{Q}_{D_{t}}) + \frac{1}{\sqrt{|\mathcal{S}|}} \left( \sqrt{\mathcal{D}_{t}(\mathcal{Q}_{D_{t}})} + \frac{1}{\sqrt{|\mathcal{S}|}} \right)$$

for $(x, y, f, \mathcal{Q}_{D_{t}}) = \max_{(x, y) \in \mathcal{S}} \left( L(x, y, f, \mathcal{Q}_{D_{t}}) < 3 \right)$, where $\mathcal{S}$ is the linear indicator vector of $f$. Proof. The argument follows [Sic].

We use the PAC-Bayesian bound
$$L(\mathcal{Q}_{D_{t}}, \mathcal{S}) = \mathbb{E}_{\mathcal{D}_{t}}[L(\mathcal{Q}_{D_{t}}, \mathcal{S})] + \frac{1}{\sqrt{|\mathcal{S}|}} \left( \sqrt{\mathcal{D}_{t}(\mathcal{Q}_{D_{t}})} + \frac{1}{\sqrt{|\mathcal{S}|}} \right)$$

where $\mathcal{D}_{t}$ denotes a prior distribution on $w_{t}$, which we set as zero-mean Gaussian, $\mathcal{D}_{t}(w_{t}) = \exp(-\frac{1}{2}||w_{t}||^{2})$. We choose $\mathcal{Q}_{D_{t}}$ as a Gaussian centered at zero, $Q_{D_{t}}(w_{t}) = \exp(-\frac{1}{2}||w_{t}||^{2})$. Then, the $L_{1}$-divergence between $Q_{D_{t}}$ and $\mathcal{D}_{t}$ is less than $3/\sqrt{2}$. Analyzing the sample risk, $\mathcal{D}_{t}(\mathcal{Q}_{D_{t}})$ can be done for any hard instance close to the sample. We denote by $L_{1}$ the predicted output for $x$ with respect to $w$. The proof is complete if we show
$$\mathbb{E}_{\mathcal{D}_{t}}[L_{1}(x, y, \mathcal{Q}_{D_{t}})] \leq \mathbb{E}_{\mathcal{D}_{t}}[L_{1}(x, y, \mathcal{S})] + \frac{1}{\sqrt{|\mathcal{S}|}}$$

where the claim follows by inequality (3) and the expression for $L_{1}(x, y, \mathcal{Q}_{D_{t}})$ using the PAC bound (3).

We show inequality (3), we use
$$\mathbb{E}_{\mathcal{D}_{t}}[L_{1}(x, y, \mathcal{Q}_{D_{t}})] = \mathbb{E}_{\mathcal{D}_{t}}[\max(0, \frac{1}{|y|}||y||^{2}||f_{t}(x, y)||_{2}^{2}||w_{t}||^{2} \leq 1)]$$

$\leq \mathbb{E}_{\mathcal{D}_{t}}[\max(0, \frac{1}{|y|}||y||^{2}||f_{t}(x, y)||_{2}^{2}||w_{t}||^{2} < 1)]$

$\leq \max(0, \frac{1}{|y|}||y||^{2}||f_{t}(x, y)||_{2}^{2}||w_{t}||^{2} > 1)$

and

$$\mathbb{E}_{\mathcal{D}_{t}}[\max(0, \frac{1}{|y|}||y||^{2}||f_{t}(x, y)||_{2}^{2}||w_{t}||^{2} \leq 1)]$$

$\leq \max(0, \frac{1}{|y|}||y||^{2}||f_{t}(x, y)||_{2}^{2}||w_{t}||^{2} > 1)$

and

$$\mathbb{E}_{\mathcal{D}_{t}}[\max(0, \frac{1}{|y|}||y||^{2}||f_{t}(x, y)||_{2}^{2}||w_{t}||^{2} > 1)]$$

$\leq \max(0, \frac{1}{|y|}||y||^{2}||f_{t}(x, y)||_{2}^{2}||w_{t}||^{2} < 1)$
What's noteworthy?

- 2000: Masters degree in Pure Mathematics
  University of Bonn, Germany
- 2003: PhD in Pure Mathematics
  University of Bonn, Germany
- 2004-2007: Postdoc in Computer Science
  DFKI Kaiserslautern, Germany
- 2006: Internship at Google
  Mountain View, California, USA
- 2007-2010: Postdoc in Computer Science
  MPI Tübingen, Germany
- 2010-2015: Assistant Professor
  ISTA, Vienna, Austria
- since 2015: Professor
  ISTA, Vienna, Austria

✔️ I completed all steps.
✔️ I switched fields after the PhD.
  - Not a problem, many scientists switch fields during their careers.

Not a good idea, experience abroad is very important!
Internships can partially compensate.
I was first Assistant Professor and later Professor at the same place.

● "tenure-track" position
What's noteworthy?

- 2000: Masters degree in Pure Mathematics  
  University of Bonn, Germany
- 2003: PhD in Pure Mathematics  
  University of Bonn, Germany
- 2004-2007: Postdoc in Computer Science  
  DFKI Kaiserslautern, Germany
- 2006: Internship at Google  
  Mountain View, California, USA
- 2007-2010: Postdoc in Computer Science  
  MPI Tübingen, Germany
- 2010-2015: Assistant Professor  
  ISTA, Vienna, Austria
- since 2015: Professor  
  ISTA, Vienna, Austria

✅ I completed all steps.
✅ I switched fields after the PhD.
  - Not a problem, many scientists switch fields during their careers.
⚠️ After the Masters, I stayed at the same university for my PhD.
  - Not a good idea, I should have moved!
What's noteworthy?

• 2000: Masters degree in Pure Mathematics
  University of Bonn, **Germany**
• 2003: PhD in Pure Mathematics
  University of Bonn, **Germany**
• 2004-2007: Postdoc in Computer Science
  DFKI Kaiserslautern, **Germany**
• 2006: Internship at Google
  Mountain View, California, USA
• 2007-2010: Postdoc in Computer Science
  MPI Tübingen, **Germany**
• 2010-2015: Assistant Professor
  ISTA, Vienna, Austria
• since 2015: Professor
  ISTA, Vienna, Austria

✅ I completed all steps.
✅ I switched fields after the PhD.
  • Not a problem, many scientists switch fields during their careers.
⚠️ After the Masters, I stayed at the same university for my PhD.
  • Not a good idea, I should have moved!
⚠️ My Masters, PhD and Postdocs were in the same country.
  • Not a good idea, experience abroad is very important!
What's noteworthy?

- **2000**: Masters degree in Pure Mathematics  
  University of Bonn, Germany

- **2003**: PhD in Pure Mathematics  
  University of Bonn, Germany

- **2004-2007**: Postdoc in Computer Science  
  DFKI Kaiserslautern, Germany

- **2006**: Internship at Google  
  Mountain View, California, USA

- **2007-2010**: Postdoc in Computer Science  
  MPI Tübingen, Germany

- **2010-2015**: Assistant Professor  
  ISTA, Vienna, Austria

- **since 2015**: Professor  
  ISTA, Vienna, Austria

- ✔️ I completed all steps.
- ✔️ I switched fields after the PhD.
  - Not a problem, many scientists switch fields during their careers.

- 🔄 After the Masters, I stayed at the same university for my PhD.
  - Not a good idea, I should have moved!
- 🔄 My Masters, PhD and Postdocs were in the same country.
  - Not a good idea, experience abroad is very important!
  - Internships can partially compensate.
What's noteworthy?

- 2000: Masters degree in Pure Mathematics
  University of Bonn, Germany
- 2003: PhD in Pure Mathematics
  University of Bonn, Germany
- 2004-2007: Postdoc in Computer Science
  DFKI Kaiserslautern, Germany
- 2006: Internship at Google
  Mountain View, California, USA
- 2007-2010: Postdoc in Computer Science
  MPI Tübingen, Germany
- 2010-2015: Assistant Professor
  ISTA, Vienna, Austria
- since 2015: Professor
  ISTA, Vienna, Austria

- ✔️ I completed all steps.
- ✔️ I switched fields after the PhD.
  - Not a problem, many scientists switch fields during their careers.
- 🔄 After the Masters, I stayed at the same university for my PhD.
  - Not a good idea, I should have moved!
- 🔄 My Masters, PhD and Postdocs were in the same country.
  - Not a good idea, experience abroad is very important!
  - Internships can partially compensate.
- ✔️ I was first Assistant Professor and later Professor at the same place.
  - “tenure-track” position
Scientific Career Steps in Academia

Standardized career path world-wide:
• Step 1: Obtain a Bachelor’s and/or Master’s Degree
• Step 2: Obtain a Doctorate/PhD
• Step 3: Work as a “Postdoc” for a few years
• Step 4: Become an Assistant Professor
• Step 5: Become a Tenured Professor

Outside of academia:
• Leave the process anywhere after Step 2

Beware: competition is fierce!
Careers of Scientists after the PhD

How to have the best chances?
How to have the best chances?

• Study at a good university.
How to have the best chances?

• Study at a good university.
How to have the best chances?

• Study at a good university. ✔

• Do your PhD at an even better institution
How to have the best chances?

• Study at a good university.

• Do your PhD at an even better institution
  – change the university, try to go abroad
  – apply to several PhD programs, but not too many: you must be able to tailor your application
    • often: some top choices, one or two “fallback” options
  – select programs that fit your interests, but don't be narrow-minded regarding topics
  – start early: up to one year between application deadline and start of the program!
Where to do a PhD?
Where to do a PhD?

Strong universities exist on every continent (except Antarctica)
• but: most university rankings target undergraduates, not PhD students

1) use resources that look at scientific publications, e.g. http://csrankings.org
• filter by research area(s) and continent/country
• scan list of faculty for potential supervisors

2) check who publishing at top venues
• don’t just look for individual big-shots, but clusters of strong people

3) find networks of excellence, e.g. ELLIS for machine learning in Europe
# CSRankings: Computer Science Rankings

[Image of a webpage showing the CSRankings interface]

<table>
<thead>
<tr>
<th>#</th>
<th>Institution</th>
<th>Count Faculty</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Carnegie Mellon University</td>
<td>25.3</td>
</tr>
<tr>
<td>2</td>
<td>Max Planck Society</td>
<td>20.2</td>
</tr>
<tr>
<td>3</td>
<td>ETH Zurich</td>
<td>19.4</td>
</tr>
<tr>
<td>4</td>
<td>Massachusetts Institute of Technology</td>
<td>15.4</td>
</tr>
<tr>
<td>5</td>
<td>Univ. of California - Berkeley</td>
<td>15.0</td>
</tr>
<tr>
<td>6</td>
<td>Stanford University</td>
<td>13.9</td>
</tr>
<tr>
<td>7</td>
<td>Cornell University</td>
<td>13.8</td>
</tr>
<tr>
<td>8</td>
<td>Univ. of Illinois at Urbana-Champaign</td>
<td>13.6</td>
</tr>
<tr>
<td>9</td>
<td>Univ. of California - San Diego</td>
<td>13.0</td>
</tr>
<tr>
<td>10</td>
<td>University of Pennsylvania</td>
<td>11.6</td>
</tr>
<tr>
<td>11</td>
<td>Tel Aviv University</td>
<td>11.3</td>
</tr>
<tr>
<td>11</td>
<td>University of Texas at Austin</td>
<td>11.3</td>
</tr>
<tr>
<td>13</td>
<td>University of Washington</td>
<td>10.4</td>
</tr>
<tr>
<td>14</td>
<td>EPFL</td>
<td>10.3</td>
</tr>
<tr>
<td>15</td>
<td>Technion</td>
<td>8.9</td>
</tr>
<tr>
<td>16</td>
<td>KAIST</td>
<td>8.8</td>
</tr>
<tr>
<td>17</td>
<td>IST Austria</td>
<td>8.4</td>
</tr>
<tr>
<td>17</td>
<td>Tsinghua University</td>
<td>8.4</td>
</tr>
<tr>
<td>19</td>
<td>University of Waterloo</td>
<td>8.1</td>
</tr>
<tr>
<td>19</td>
<td>University of Wisconsin - Madison</td>
<td>8.1</td>
</tr>
<tr>
<td>#</td>
<td>Institution</td>
<td>Count Faculty</td>
</tr>
<tr>
<td>----</td>
<td>----------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>1</td>
<td>Max Planck Society</td>
<td>20.2</td>
</tr>
<tr>
<td>2</td>
<td>ETH Zurich</td>
<td>19.4</td>
</tr>
<tr>
<td>3</td>
<td>Tel Aviv University</td>
<td>11.3</td>
</tr>
<tr>
<td>4</td>
<td>EPFL</td>
<td>10.3</td>
</tr>
<tr>
<td>5</td>
<td>Technion</td>
<td>8.9</td>
</tr>
<tr>
<td>6</td>
<td>IST Austria</td>
<td>8.4</td>
</tr>
<tr>
<td>7</td>
<td>University of Edinburgh</td>
<td>6.7</td>
</tr>
<tr>
<td>8</td>
<td>Ecole Normale Superieure</td>
<td>6.6</td>
</tr>
<tr>
<td>9</td>
<td>TU Munich</td>
<td>6.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Faculty</th>
<th># Pubs</th>
<th>Adj. #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Krishnendu Chatterjee</td>
<td>LOGIC</td>
<td>14.5</td>
</tr>
<tr>
<td>Bernd Bickel</td>
<td>GRAPHICS</td>
<td>6.6</td>
</tr>
<tr>
<td>Christoph H. Lampert</td>
<td>VISION,ML</td>
<td>11.6</td>
</tr>
<tr>
<td>Thomas A. Henzinger</td>
<td>LOGIC</td>
<td>7.0</td>
</tr>
<tr>
<td>Dan Alistarh</td>
<td>ML</td>
<td>5.4</td>
</tr>
<tr>
<td>Christopher Wojtan</td>
<td>GRAPHICS</td>
<td>6.2</td>
</tr>
<tr>
<td>Krzysztof Pietrzak</td>
<td>CRYPTO</td>
<td>4.7</td>
</tr>
<tr>
<td>Vladimir Kolmogorov</td>
<td>VISION,THEORY</td>
<td>3.2</td>
</tr>
<tr>
<td>Marco Mondelli</td>
<td>ML</td>
<td>2.5</td>
</tr>
<tr>
<td>Gasper Tkacik</td>
<td>ML</td>
<td>0.8</td>
</tr>
<tr>
<td>Caelin C. Guet</td>
<td></td>
<td>0.2</td>
</tr>
</tbody>
</table>
ELLIS - the European Laboratory for Learning and Intelligent Systems - is a pan-European AI network of excellence which focuses on fundamental science, technical innovation and societal impact. Founded in 2018, ELLIS builds upon machine learning as the driver for modern AI and aims to secure Europe's sovereignty in this competitive field by creating a multi-centric AI research laboratory. ELLIS wants to ensure that the highest level of AI research is performed in the open societies of Europe and follows a three-pillar strategy to achieve that.
ELLIS Units (as of 2022)

- Alicante
- Amsterdam
- Copenhagen
- Darmstadt
- Lausanne (EPFL)
- Zürich (ETH)
- Haifa (Technion)
- Heidelberg
- Vienna (IST Austria)
- Leuven
- London (UCL)
- Madrid
- Modena (Unimore)
- Munich
- Paris
- Prague
- Tel Aviv
- Tübingen
- Berlin
- Cambridge
- Delft
- Edinburgh
- Freiburg
- Genoa
- Helsinki
- Jena
- Linz
- Lisbon
- Manchester
- Milan
- Nijmegen
- Oxford
- Saarbrücken
- Stuttgart
- Turin
How to find a good supervisor?

1) **scientific quality matters**
   - excellent scientists are not automatically excellent supervisors
   - but mediocre scientists will not be able to make you excellent

2) **the past is the best predictor of the future:**
   - check out who graduated from the potential supervisors' groups over the last year
   - what career path did they take? were they successful? would you like to end up like them?

3) **ask (also) the group members:**
   - supervisors are overwhelmed with email, they might not reply, or only superficially
   - group members a) have more time, b) are more open if the supervision is good or not
What to look for in a PhD program?

Running example:
the PhD Program at ISTA
Check the format

• Traditional European Master-Apprentice system
  – supervisor hires PhD student
  – usually requires Masters degree
  – few additional support structure

• (US-style) graduate school system
  – centralized admissions process
  – sometimes: initial ‘unaffiliated’ phase with courses and/or rotation projects
  – enter with a BSc or MSc degree
  – for BSc entry, possible en-route MSc degree
Example: ISTA PhD program

Phase I

Core Requirement
- Core Project
- Track Core Course
- Essential Skills for Scientists

Elective Requirement
- 12 or 24 ECTS of coursework
- Can also be (partially) completed in Phase II

Optional Service Courses
- Skills courses in programming, mathematics and lab techniques

Rotation 1 Rotation 2 Rotation 3
(optimal)
Rotation 4 Rotation 5

Affiliation with Research Group Qualifying Exam

Phase II

PhD Thesis Research

Teaching (as TA)

Regular Progress Reviews

Annual Research Presentations

Thesis Defense

PhD Degree
Gradschool is more than a job

• Research requires personal connections
  – supervisor is boss but also mentor
  – research groups offer team experience
  – collaborators and thesis committee provide external advice

• Strong feeling of coherence between students
  – life-long connections across discipline boundaries
  – often cross-cultural experiences
    e.g. ISTA: 280 students from 54 countries
Gradschool is also a job

Check for fair treatment of PhD students:
- competitive salary (full position), social security?
- are PhD students supported, e.g. with a travel budget?
- additional benefits?
  - on-campus housing?
  - public transport?
- campus life?
  - e.g. ISTA: soccer field, tennis courts, volleyball court, in-house gym, restaurant, bar, kindergarten, ...
Last but not least: location!

In the institution located in “The murder capital of country X”? or in “The most livable city worldwide” *?

* That’s Vienna, according to Mercer’s Quality of Living Rankings
Getting into a strong PhD program

Running example:
the PhD Program at ISTA
Getting into a strong PhD program

• Are good graduate schools hard to get in? Absolutely...

• ISTA PhD program: acceptance rate ~4% in 2021
  – over 4000 interested applicants
  – 2569 submitted applications
  – ~200 interviews (online and on-campus)
  – 106 offers
  – 67 accepted
Getting into a strong PhD program

How to maximize your chances?

• Prepare:
  – a lot of material is online, check out the program websites
  – identify potential programs, ideally more than one
  – identify potential supervisors, ideally more than one

• Try to stand out from the crowd:
  – talk to potential supervisors at workshops/meetings
    • you can try email, but often that’s too anonymous
  – consider doing an internship before applying:
    e.g. http://ist.ac.at/research/internships/
Getting into a strong PhD program

Requirements: Bachelor's or Master's degree, depending on the program
(by the time the program starts)

Application material:
• 1) resume
• 2) transcripts of your BS and/or MS degree
• 3) statement of purpose
• 4) contact details of three referees
  – reference letters will be uploaded by the referees, not by you
• 5) some places: English language certificates (e.g. TOEFL)
Getting into a strong PhD program

Resume

- tabular academic resume:
  - usually 1—2 pages
  - English language
  - inverse chronological order
  - no (truly) personal data required:
    - photo, marital status, religion, hobbies, ...
  - emphasize education over work experience
  - include relevant experiences/achievements:
    - awards, internships, publications, language skills, ...
Getting into a strong PhD program

Resume
Getting into a strong PhD program

**Transcripts**

- transcripts of Bachelor and Master degree (if available)
  - courses taken
  - grades
  - if grading system is complicated: provide explanation
  - if not in English: provide translation
    - some places ask for certified translations, ISTA does not

Note: Master’s/Bachelor’s grades do matter! PhD grades don’t.
Getting into a strong PhD program

**Statement of Purpose**

- between 1 and 2 pages:
  - why do you want to do a PhD?
  - why at this institution?
  - what research are you interested in?
  - also: opportunity to explain things that might be awkward in the other documents
    - bad grades, gaps in the CV, ...
- be honest, but don't be modest
Getting into a strong PhD program

Statement of Purpose

looks good
too short
Getting into a strong PhD program

Statement of Purpose

much too long (complete PhD topic proposal)
Getting into a strong PhD program

Referees

(reference letters are surprisingly important, choose well)

Note: most good places will contact referees directly for letters. If you attach any letters yourself, they will be ignored.

• most important: reference letters must be positive and strong
  – not “She's an okay student.”
  – rather “She's the smartest student I ever met.”

• also important: reference letters must be personal
  – not “I don't really know him well.”
  – rather “I supervised his master thesis.”
Getting into a strong PhD program

Referees

(reference letters are surprisingly important, choose well)

• also important: **referees should know you scientifically**
  – not “I'm her soccer coach.” or “I'm his brother.”
  – rather “She did an internship with me for six months.”

• also: **try to choose a diverse set of referees**
  – not three course teachers from the same university where you study
  – ideally: different countries, or at least different institutions

• also: **scientific reputation of referees matters** as well
  – avoid: graduate student or first year postdoc
  – preferable: internationally well-known professor at top university
  – But: when in doubt, choose strong and personal reference over lukewarm one from a big-shot

• **Look for potential referees already before you need them!**
Getting into a strong PhD program

**Publications?**

• if you have any, list them

• **first-author publications at top-tier venues** are taken as sign of excellence
  – for some competitive CS programs, such as UC Berkeley, almost a requirement

• **other publications** (low-tier, or as middle author) demonstrate that at least you participated in research work and experienced the process
  – the acquired soft skills will count as positive
  – but: the contents of the publication will matter little
Getting into a strong PhD program

Start early

- Application portals open in late autumn
  - Oct 1
- Deadline to apply and submit: winter
  - Nov 15
- Interviews (selected candidates): early spring
  - Jan/Feb
- Admissions offers: soon afterwards
  - Feb/Mar
- Deadline to accept or decline offers: usually April 15
  - Apr 15
- Programs start: next autumn
  - Sep/Oct

How to fill the summer gap?

- summer schools, internships, vacation, volunteer work..
Behind the scenes: Life of a (Machine Learning) Researcher
How do researchers spend their time?

Source: non-representative survey from Boise State University: https://thebluereview.org/faculty-time-allocation/
How do researchers spend their time?

**Administration**
- meetings
  - research group
  - institution
  - project teams
- recruiting
- reporting

**Research**
- actual real research
- publications:
  - writing
  - reading
  - reviewing
- grant proposals:
  - writing
  - reviewing

**Education**
- lectures
  - preparation
  - teaching
  - grading exams/homeworks
- supervision
- mentoring
<table>
<thead>
<tr>
<th>Administration</th>
<th>Research</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>• meetings</td>
<td>• actual real research</td>
<td>• lectures</td>
</tr>
<tr>
<td>- research group</td>
<td>• publications:</td>
<td>– preparation</td>
</tr>
<tr>
<td>- institution</td>
<td>- writing</td>
<td>– teaching</td>
</tr>
<tr>
<td>- project teams</td>
<td>- reading</td>
<td>– grading exams/</td>
</tr>
<tr>
<td>• recruiting</td>
<td>- reviewing</td>
<td>homeworks</td>
</tr>
<tr>
<td>• reporting</td>
<td>• grant proposals:</td>
<td>• supervision</td>
</tr>
<tr>
<td></td>
<td>- writing</td>
<td>• mentoring</td>
</tr>
<tr>
<td></td>
<td>- reviewing</td>
<td></td>
</tr>
</tbody>
</table>

**PhD student**
How do researchers spend their time?

**Postdoc in academia**

**Administration**
- meetings
  - research group
  - institution
  - project teams
- recruiting
- reporting

**Research**
- actual real research
- publications:
  - writing
  - reading
  - reviewing
- grant proposals:
  - writing
  - reviewing

**Education**
- lectures
  - preparation
  - teaching
  - grading exams/homeworks
- supervision
- mentoring

**PhD student**
How do researchers spend their time?

**Researcher in industry**
- **Administration**
  - meetings
    - research group
    - institution
    - project teams
  - recruiting
  - reporting
- **Research**
  - actual real research
  - publications:
    - writing
    - reading
    - reviewing
  - grant proposals:
    - writing
    - reviewing

**Postdoc in academia**
- **Education**
  - lectures
    - preparation
    - teaching
    - grading exams/homeworks
  - supervision
  - mentoring
- **Administration**
  - meetings
    - research group
    - institution
    - project teams
  - recruiting
  - reporting
How do researchers spend their time?

**Researcher in industry**
- **Administration**
  - meetings
    - research group
    - institution
    - project teams
  - recruiting
  - reporting
- **Research**
  - actual real research
  - publications:
    - writing
    - reading
    - reviewing
  - grant proposals:
    - writing
    - reviewing

**Postdoc in academia**
- **Education**
  - lectures
    - preparation
    - teaching
    - grading exams/homeworks
  - supervision
  - mentoring

**PhD student**

**Faculty in academia**
How do researchers spend their time?

**Researcher in industry**
- Administration
  - meetings
    - research group
    - institution
    - project teams
  - recruiting
  - reporting
- Research
  - actual real research
  - publications:
    - writing
    - reading
    - reviewing
  - grant proposals:
    - writing
    - reviewing

**Postdoc in academia**
- Education
  - lectures
    - preparation
    - teaching
    - grading exams/homeworks
  - supervision
  - mentoring

**Faculty in academia**
- Administration
  - meetings
    - research group
    - institution
    - project teams
  - recruiting
  - reporting

50%
How do researchers spend their time?

**Researcher in industry**
- Administration
  - meetings
    - research group
    - institution
    - project teams
  - recruiting
  - reporting

**Research**
- actual real research
- publications:
  - writing
  - reading
  - reviewing
- grant proposals:
  - writing
  - reviewing

**Postdoc in academia**
- Education
  - lectures
    - preparation
    - teaching
    - grading exams/homeworks
  - supervision
  - mentoring

**PhD student**
- lectures
  - preparation
  - teaching
  - grading exams/homeworks
  - supervision
  - mentoring

**Faculty in academia**

50%  

50%
How do researchers spend their time?

**Researcher in industry**
- Administration
  - meetings
    - research group
    - institution
    - project teams
  - recruiting
  - reporting

**Research**
- actual real research
- publications:
  - writing
  - reading
  - reviewing
- grant proposals:
  - writing
  - reviewing

**Postdoc in academia**
- Education
  - lectures
    - preparation
    - teaching
    - grading exams/homeworks
  - supervision
  - mentoring

**PhD student**

**Faculty in academia**

50% 50% 50%
Behind the scenes: Publish or Perish
Behind the scenes: Publish or Perish

- more than 2.5 million new publications per year
- ~5% (150,000) in artificial intelligence/robotics
- clearly growing trend: 35% increase 2011-2019

Behind the scenes: Publish or Perish

In a large research area, such as machine learning, it's impossible to stay up-to-date with all works!

One has to rely on other mechanisms to identify what papers (or at least their titles) to read:

<table>
<thead>
<tr>
<th></th>
<th>Peer Review</th>
<th>Name Recognition</th>
<th>Aggregators</th>
<th>Social Media</th>
</tr>
</thead>
<tbody>
<tr>
<td>what to read?</td>
<td>what is published at top conferences or journals</td>
<td>publications from top research labs</td>
<td>automatic digests, e.g. Google Scholar, based on keywords or citations</td>
<td>what shows up on Twitter or Youtube</td>
</tr>
<tr>
<td>how much?</td>
<td>a few thousand papers every couple of weeks</td>
<td>a few dozen papers every couple of weeks</td>
<td>100 papers per week (adjustable)</td>
<td>10 papers per day</td>
</tr>
<tr>
<td>problems?</td>
<td>still too much; mainstream bias: what’s currently trendy in the community</td>
<td>too little; rich-get-richer; narrow coverage</td>
<td>filter bubble; focus on arXiv preprints</td>
<td>hype-driven; filter bubble; focus on arXiv preprints</td>
</tr>
</tbody>
</table>

Advice to young scientist: To be read, you have to make oneself visible: scientific homepage, Google Scholar profile, upload manuscripts to arXiv, Twitter account, socialize...
The long path of a (machine learning) publication

- original research
- manuscript is written
- manuscript is submitted to conference/workshop
- meta-reviewer(s) select reviewers from a pool of "volunteers"
- manuscript is reviewed (typically by 3 reviewers)
- meta-reviewers chase late reviewers
- some top venues

- manuscripts uploaded to arXiv.org
- desk reject (formal reasons)

- authors receive reviews
- authors address reviews (in author response or revision)

- reviewers discuss & update reviews
- meta-reviewer(s) make(s) decision
- authors prepare camera-ready
- conference chairs confirm decisions
- reject (60-80%)

- authors present at conference as poster or talk
- papers is officially published

- manuscripts is re-written
- most conferences/workshops
- authors receive reviews
The long path of a (machine learning) publication

1. **Original research** (a few months)
   - **Manuscript is written** (a few weeks)

2. **Manuscript is submitted to conference/workshop** (hard deadline)
   - **Meta-reviewer(s) select reviewers** (from a pool of "volunteers")
   - **Desk reject** (formal reasons)
     - **Manuscript is rejected** (50-80%)
     - **Manuscript is re-written** (hours to months)

3. **Manuscript is re-written**
   - **Meta-reviewer(s) make(s) decision**
     - **Reviewers discuss & update reviews**
     - **Authors address reviews** (in author response or revision)
     - **Authors receive reviews**
     - **Conference chairs confirm decisions**
     - **Authors prepare camera-ready**

4. **Authors present at conference as poster or talk**
   - **Conference chairs confirm decisions**
   - **Manuscript is officially published**
     - **Papers is officially published**
     - **Authors address reviews**
   - **Manuscript is uploaded to arXiv.org** (2 weeks)

5. **Some top venues**
   - **Meta-reviewers chase late reviewers** (2 more weeks)
   - **Authors present at conference as poster or talk**

6. **A few months**
   - **A few weeks**

---

- **Accept**
- **Reject** (50-80%)
- **Meta-reviewer(s) make(s) decision**
- **Reviewers discuss & update reviews**
- **Authors address reviews** (in author response or revision)
- **Authors receive reviews**
- **Conference chairs confirm decisions**
- **Manuscript is officially published**
- **Papers is officially published**
- **Manuscript is uploaded to arXiv.org**
Behind the scenes: Publish or Perish

ICLR 2021 Ratings Averaged By Paper

Source: https://github.com/evanzd/ICLR2021-OpenReviewData
Decision boundary: approximately 6.0 (though with substantial overlap) → some luck is required, too.

Source: https://github.com/evanzd/ICLR2021-OpenReviewData
How much luck?

Consistency experiments at NeurIPS 2021:
• 8765 submissions (overall acceptance rate: ~22%)
• for 882 (10%) a second copy was processed twice independently

<table>
<thead>
<tr>
<th></th>
<th>2nd Accept</th>
<th>2nd Reject</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Accept</td>
<td>99</td>
<td>107</td>
</tr>
<tr>
<td>1st Reject</td>
<td>96</td>
<td>462</td>
</tr>
</tbody>
</table>

Analysis:
• fraction of inconsistent decisions: 23%.
• if acceptance were purely random: 35%

Summary:
• luck matters, but scientific quality does as well
• results consistent with similar analysis in 2014 → at least, the process is not getting worse

Source: https://blog.neurips.cc/2021/12/08/the-neurips-2021-consistency-experiment/
Behind the scenes: Research Grants
Who pays for all this?
World-wide Research Spending
World-wide Research Spending

>2 trillion USD (~2.6% of world GDP)

Research Funding

Most academic researchers rely on competitive research grants to fund their (group’s) research:
• national funding organizations
• European funding organizations
• world-wide programs, e.g. philantropies

Some numbers (somewhat anecdotal):
• group leaders on average spend 40% of their time writing grant applications
• success rate: usually below 20%
• cost to prepare proposals: 15% of call budget

Expenditure of research and development in Estonia, 2019 (in MEUR)

Source: https://researchinestonia.eu
Competitive Research Grants

Individual/stand-alone grant:
• generally: funds (partially) one PhD student or postdoc working on a specific project
• rarely (e.g. ERC Grants): funds a research team at a single institution

Collaborative grant:
• funds a team working on a specific project, distributed across multiple institutions

Excellence Cluster/Network of Excellence/Doctoral School:
• many positions at single or multiple institutions to work on different (related) projects

... and many many others
Collaborative Grants

Example: European Union’s “Horizon Europe“ – total budget: over 95 billion EUR

- typically calls about specific research directions, e.g. “green and sustainable innovation”

Before the project:

- three to eight partners (or more) from at least three different countries form a consortium
- prepare a joint proposal: project idea, prior work, solution path, cost breakdown, researcher resumes and publication lists, information on hosting institution, planned outreach, ethics forms, endorsement letters from industry partners...

During the project:

- time-sheets: who worked on which parts of the project for how long
- intermediate reports about milestones and deliverables
- intermediate review meetings with external experts

After the project:

- write a final report: scientific outcome, development of human resources, impact beyond the project itself, efficiency of resource usage, perspectives of future possibilities
Behind the scenes:
You’re not in it alone – collaboration networks
Research is a collaborative effort

A network of collaborators allows:

- joint research projects,
- joint paper writing,
- joint grant proposals,
- joint event organization,
- PhD committee memberships,
- recommendation letters,
- exchange of people,
- dissemination of scientific results,
- ... more fun at conferences and workshops.

It’s never too early to start a network of (future) collaborators!