ESSCaSS 2022

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

Christoph Lampert

ISTA Institute of Science and Technology Austria

Publicly-Funded Research Institute

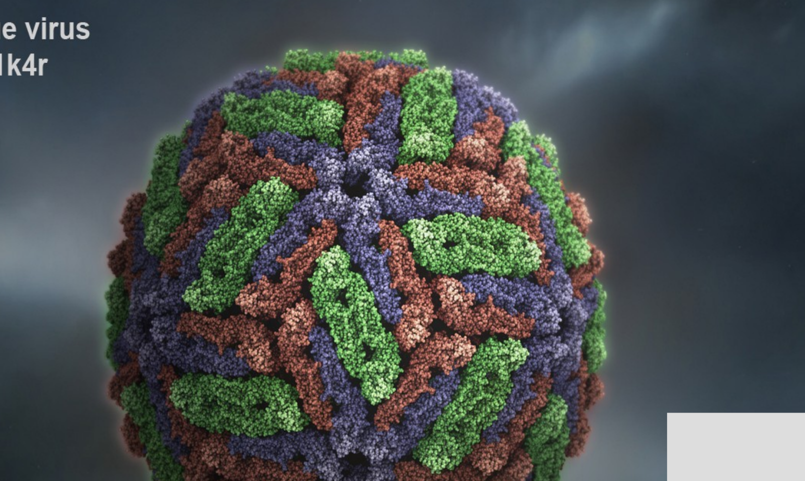
- PhD-granting graduate school
- no undergraduate studies (but internships)
- founded in 2009
- located close to Vienna, Austria

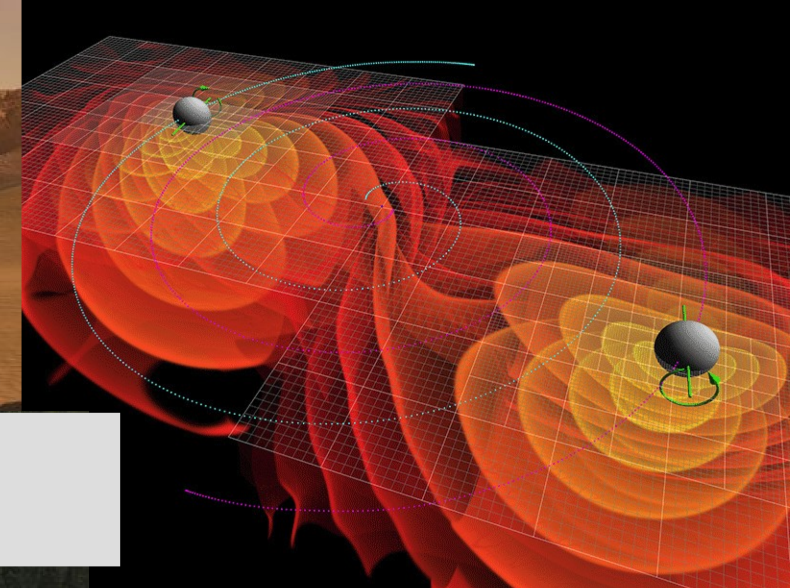
Focus on

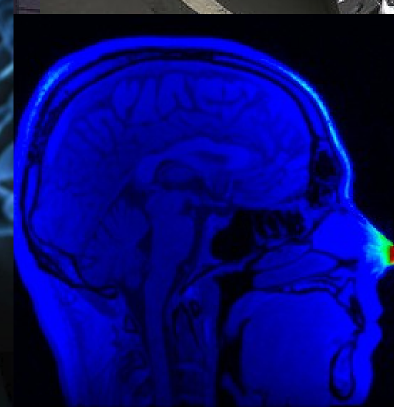
- curiosity-driven basic research
- interdisciplinarity: Computer Science, Mathematics, Biology, Physics, Chemistry, Earth&Climate Sciences, Neuroscience

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.**

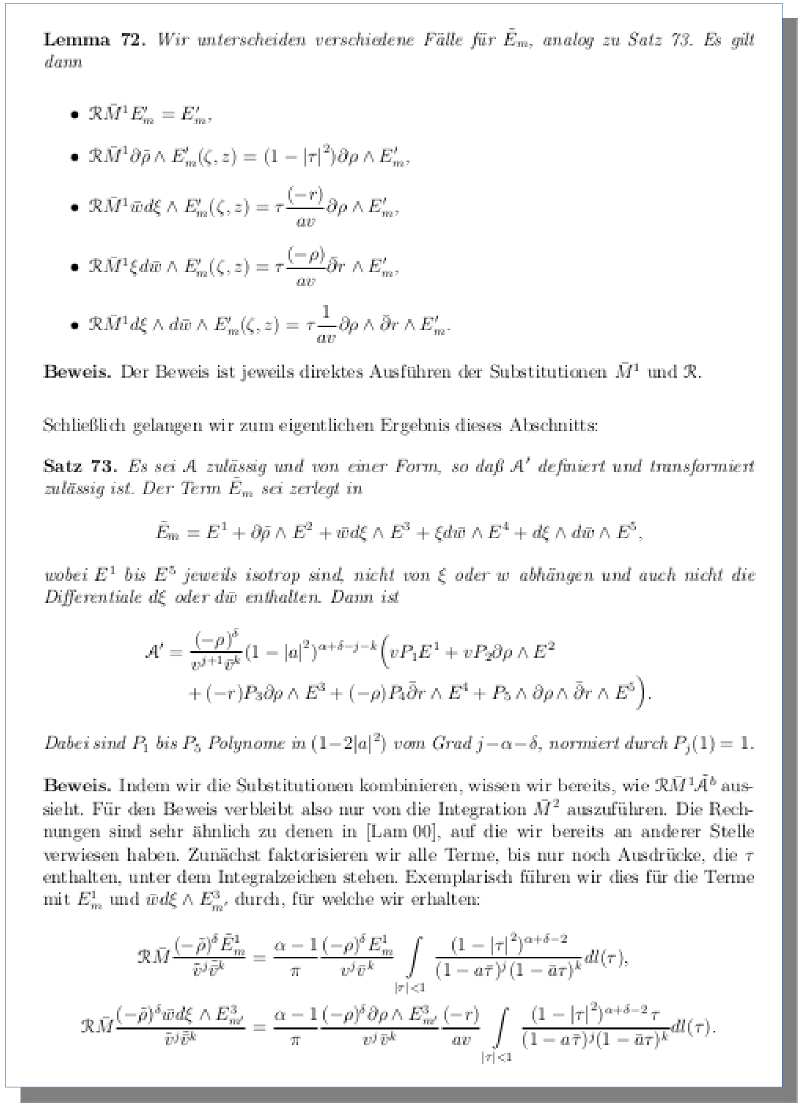
✓ **I switched fields after the PhD.**

before:

Complex Analysis

after:

Computer Vision



Lemma 72. *Wir unterscheiden verschiedene Fälle für $\bar{E}_m$, analog zu Satz 73. Es gilt dann*

- $\mathcal{R}\bar{M}^1 E'_m = E'_m$,
- $\mathcal{R}\bar{M}^1 \partial\bar{\rho} \wedge E'_m(\zeta, z) = (1-|\tau|^2)\partial\rho \wedge E'_m$,
- $\mathcal{R}\bar{M}^1 \bar{w} d\xi \wedge E'_m(\zeta, z) = \tau\frac{(-r)}{av}\partial\rho \wedge E'_m$,
- $\mathcal{R}\bar{M}^1 \xi d\bar{w} \wedge E'_m(\zeta, z) = \tau\frac{(-\rho)}{av}\bar{\partial}r \wedge E'_m$,
- $\mathcal{R}\bar{M}^1 d\xi \wedge d\bar{w} \wedge E'_m(\zeta, z) = \tau\frac{1}{av}\partial\rho \wedge \bar{\partial}r \wedge E'_m$.

Beweis. Der Beweis ist jeweils direktes Ausführen der Substitutionen $\bar{M}^1$ und $\mathcal{R}$.

Schließlich gelangen wir zum eigentlichen Ergebnis dieses Abschnitts:

Satz 73. *Es sei $\mathcal{A}$ zulässig und von einer Form, so daß $\mathcal{A}'$ definiert und transformiert zulässig ist. Der Term $\bar{E}_m$ sei zerlegt in*

$$\bar{E}_m = E^1 + \partial\bar{\rho} \wedge E^2 + \bar{w} d\xi \wedge E^3 + \xi d\bar{w} \wedge E^4 + d\xi \wedge d\bar{w} \wedge E^5,$$

wobei E^1 bis E^5 jeweils isotrop sind, nicht von ξ oder w abhängen und auch nicht die Differentiale $d\xi$ oder $d\bar{w}$ enthalten. Dann ist

$$\mathcal{A}' = \frac{(-\rho)^\delta}{v^{j+1}\bar{v}^k}(1-|a|^2)^{\alpha+\delta-j-k}\Big(vP_1E^1 + vP_2\partial\rho \wedge E^2$$
$$+ (-r)P_3\partial\rho \wedge E^3 + (-\rho)P_4\bar{\partial}r \wedge E^4 + P_5 \wedge \partial\rho \wedge \bar{\partial}r \wedge E^5\Big).$$

Dabei sind P_1 bis P_5 Polynome in $(1-2|a|^2)$ vom Grad $j-\alpha-\delta$, normiert durch $P_j(1) = 1$.

Beweis. Indem wir die Substitutionen kombinieren, wissen wir bereits, wie $\mathcal{R}\bar{M}^1\bar{\mathcal{A}}^b$ aussieht. Für den Beweis verbleibt also nur von die Integration $\bar{M}^2$ auszuführen. Die Rechnungen sind sehr ähnlich zu denen in [Lam 00], auf die wir bereits an anderer Stelle verwiesen haben. Zunächst faktorisieren wir alle Terme, bis nur noch Ausdrücke, die τ enthalten, unter dem Integralzeichen stehen. Exemplarisch führen wir dies für die Terme mit E^1_m und $\bar{w}d\xi \wedge E^3_{m'}$ durch, für welche wir erhalten:

$$\mathcal{R}\bar{M}\frac{(-\bar{\rho})^\delta \bar{E}^1_m}{\bar{v}^j\bar{\bar{v}}^k} = \frac{\alpha-1}{\pi}\frac{(-\rho)^\delta E^1_m}{v^j\bar{v}^k}\int\limits_{|\tau|<1}\frac{(1-|\tau|^2)^{\alpha+\delta-2}}{(1-a\bar{\tau})^j(1-\bar{a}\tau)^k}dl(\tau),$$

$$\mathcal{R}\bar{M}\frac{(-\bar{\rho})^\delta \bar{w}d\xi \wedge E^3_{m'}}{\bar{v}^j\bar{\bar{v}}^k} = \frac{\alpha-1}{\pi}\frac{(-\rho)^\delta \partial\rho \wedge E^3_{m'}}{v^j\bar{v}^k}\frac{(-r)}{av}\int\limits_{|\tau|<1}\frac{(1-|\tau|^2)^{\alpha+\delta-2}\tau}{(1-a\bar{\tau})^j(1-\bar{a}\tau)^k}dl(\tau).$$

Figure 6. Quality of individual attribute predictors (trained on *train* classes, tested on *test* classes), as measured by by *area under ROC curve (AUC)*. Attributes with zero entries have constant values for all test classes of the split chosen, so their AUC is undefined.

Figure 7. Highest ranking results for each test class in the *Animals with Attributes* dataset. Uniquely characterized classes are identified well, *e.g.* humpback whales and leopards. Confusions occur between visually similar categories, *e.g.* pigs and hippopotamuses.

ages for training and testing. Clearly, the availability of a large number of training samples from the same classes as the test data vastly simplifies the problem. With a resulting multi-class accuracy of 65.9%, supervised training does indeed perform better than the 40.5% achieved by *attribute-based learning*. However, given the different amount of training information included, we believe that the difference in accuracy is not discouragingly large, and that learning with attributes has the potential to complement supervised classification in areas where no or only few training examples are available.

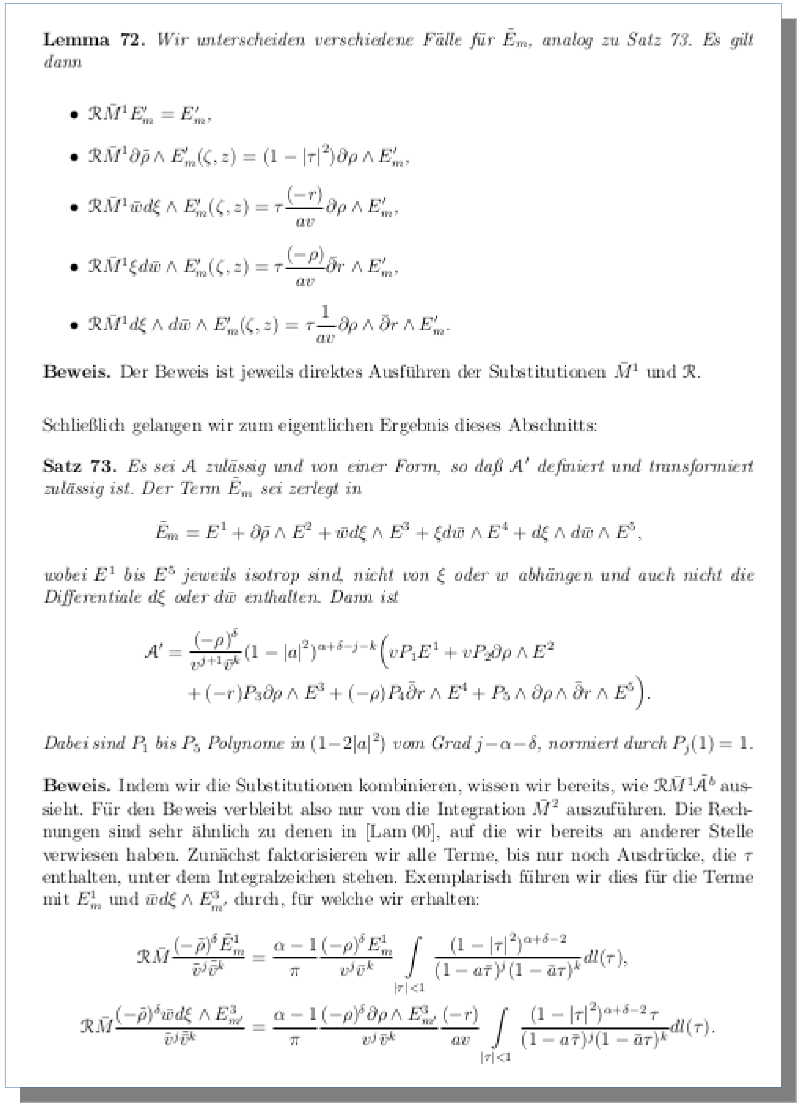
6. Conclusion

In this paper, we have introduced *learning for disjoint training and test classes*. It formalizes the problem of learning an object classification systems for classes, for which no training images are available. We have proposed two methods for *attribute-based classification* that solve this problem by transferring information between classes. The transfer is achieved by an intermediate representation that consists of high level, semantic, per-class attributes, providing a fast and easy way to include human knowledge into the system. Once trained, the system can detect any object category, for which a suitable characterization by attributes is available, and it does not require a re-training step.

Additionally, we have introduced the *Animals with Attributes* dataset: it consists over 30,000 images with precomputed reference features for 50 animal classes, for which a semantic attribute annotation is available from studies in cognitive science. We hope that this dataset will facilitate research and serve as a testbed for *attribute-based classification*.

Starting from the proposed system, many improvements and extensions are possible. Clearly, better designed per-attribute and multi-class classifiers could improve the overall performance of the system, as could a per-attribute feature selection set, because clearly not all attributes relevant for a human can be determined from images. For an adaptive system that can grow to include new classes, it should be possible to increase the attribute set without retraining. It would also be very interesting to remove the amount of human effort, *e.g.* by letting a human define the attributes, but build an automatic system to label the image classes with the attribute values, possibly using textual information from before:

Complex Analysis



Lemma 72. *Wir unterscheiden verschiedene Fälle für $\bar{E}_m$, analog zu Satz 73. Es gilt dann*

- $\mathcal{R}\bar{M}^1 E'_m = E'_m$,
- $\mathcal{R}\bar{M}^1 \partial\bar{\rho} \wedge E'_m(\zeta, z) = (1 - |\tau|^2)\partial\rho \wedge E'_m$,
- $\mathcal{R}\bar{M}^1 \bar{w} d\xi \wedge E'_m(\zeta, z) = \tau \frac{(-r)}{av} \partial\rho \wedge E'_m$,
- $\mathcal{R}\bar{M}^1 \xi d\bar{w} \wedge E'_m(\zeta, z) = \tau \frac{(-\rho)}{av} \bar{\partial} r \wedge E'_m$,
- $\mathcal{R}\bar{M}^1 d\xi \wedge d\bar{w} \wedge E'_m(\zeta, z) = \tau \frac{1}{av} \partial\rho \wedge \bar{\partial} r \wedge E'_m$.

Beweis. Der Beweis ist jeweils direktes Ausführen der Substitutionen $\bar{M}^1$ und $\mathcal{R}$.

Schließlich gelangen wir zum eigentlichen Ergebnis dieses Abschnitts:

Satz 73. *Es sei $\mathcal{A}$ zulässig und von einer Form, so daß $\mathcal{A}'$ definiert und transformiert zulässig ist. Der Term $\bar{E}_m$ sei zerlegt in*

$$\bar{E}_m = E^1 + \partial\bar{\rho} \wedge E^2 + \bar{w} d\xi \wedge E^3 + \xi d\bar{w} \wedge E^4 + d\xi \wedge d\bar{w} \wedge E^5,$$

wobei E^1 bis E^5 jeweils isotrop sind, nicht von ξ oder w abhängen und auch nicht die Differentiale $d\xi$ oder $d\bar{w}$ enthalten. Dann ist

$$\mathcal{A}' = \frac{(-\rho)^\delta}{v^{j+1}\bar{v}^k}(1 - |a|^2)^{\alpha+\delta-j-k}\Big(vP_1E^1 + vP_2\partial\rho \wedge E^2 + (-r)P_3\partial\rho \wedge E^3 + (-\rho)P_4\bar{\partial} r \wedge E^4 + P_5 \wedge \partial\rho \wedge \bar{\partial} r \wedge E^5\Big).$$

Dabei sind P_1 bis P_5 Polynome in $(1-2|a|^2)$ vom Grad $j-\alpha-\delta$, normiert durch $P_j(1) = 1$.

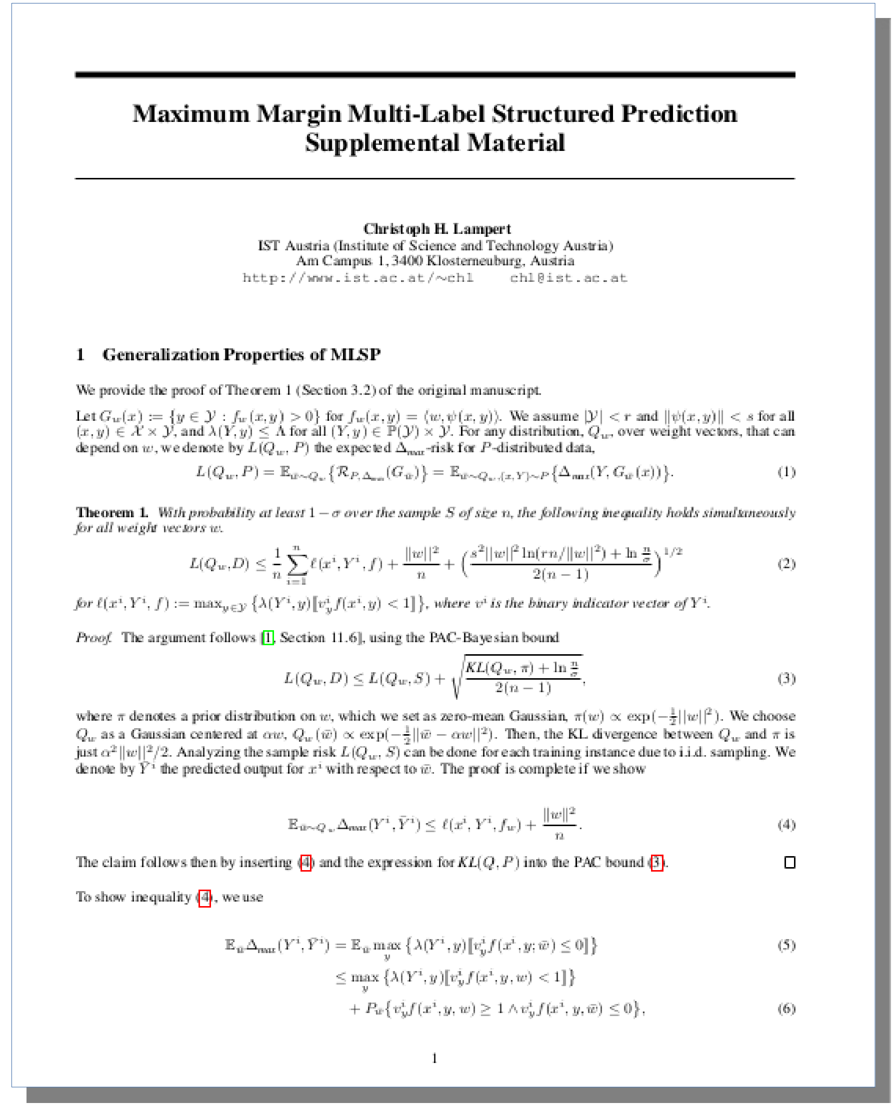
Beweis. Indem wir die Substitutionen kombinieren, wissen wir bereits, wie $\mathcal{R}\bar{M}^1\bar{\mathcal{A}}^b$ aussieht. Für den Beweis verbleibt also nur von die Integration $\bar{M}^2$ auszuführen. Die Rechnungen sind sehr ähnlich zu denen in [Lam 00], auf die wir bereits an anderer Stelle verwiesen haben. Zunächst faktorisieren wir alle Terme, bis nur noch Ausdrücke, die τ enthalten, unter dem Integralzeichen stehen. Exemplarisch führen wir dies für die Terme mit E^1_m und $\bar{w} d\xi \wedge E^3_{m'}$ durch, für welche wir erhalten:

$$\mathcal{R}\bar{M}\frac{(-\bar{\rho})^\delta \bar{E}^1_m}{\bar{v}^j\bar{\bar{v}}^k} = \frac{\alpha - 1}{\pi}\frac{(-\rho)^\delta E^1_m}{v^j\bar{v}^k}\int\limits_{|\tau|<1}\frac{(1-|\tau|^2)^{\alpha+\delta-2}}{(1-a\bar{\tau})^j(1-\bar{a}\tau)^k}dl(\tau),$$

$$\mathcal{R}\bar{M}\frac{(-\bar{\rho})^\delta \bar{w} d\xi \wedge E^3_{m'}}{\bar{v}^j\bar{\bar{v}}^k} = \frac{\alpha - 1}{\pi}\frac{(-\rho)^\delta \partial\rho \wedge E^3_{m'}}{v^j\bar{v}^k}\frac{(-r)}{av}\int\limits_{|\tau|<1}\frac{(1-|\tau|^2)^{\alpha+\delta-2}\tau}{(1-a\bar{\tau})^j(1-\bar{a}\tau)^k}dl(\tau).$$

also after:

Machine Learning



Maximum Margin Multi-Label Structured Prediction
Supplemental Material

Christoph H. Lampert
IST Austria (Institute of Science and Technology Austria)
Am Campus 1, 3400 Klosterneuburg, Austria
http://www.ist.ac.at/~chl   chl@ist.ac.at

1 Generalization Properties of MLSP

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

Let $G_w(x) := \{y \in \mathcal{Y} : f_w(x,y) > 0\}$ for $f_w(x,y) = \langle w, \psi(x,y)\rangle$. We assume $|\mathcal{Y}| < r$ and $\|\psi(x,y)\| < s$ for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$, and $\lambda(Y,y) \leq \Lambda$ for all $(Y,y) \in \mathbb{P}(\mathcal{Y}) \times \mathcal{Y}$. For any distribution, Q_w, over weight vectors, that can depend on w, we denote by $L(Q_w, P)$ the expected Δ_{max}-risk for P-distributed data,

$$L(Q_w, P) = \mathbb{E}_{\bar{w}\sim Q_w}\{\mathcal{R}_{P,\Delta_{max}}(G_{\bar{w}})\} = \mathbb{E}_{\bar{w}\sim Q_w,(x,Y)\sim P}\{\Delta_{max}(Y, G_{\bar{w}}(x))\}. \quad (1)$$

Theorem 1. *With probability at least $1-\sigma$ over the sample S of size n, the following inequality holds simultaneously for all weight vectors w.*

$$L(Q_w,D) \leq \frac{1}{n}\sum_{i=1}^n \ell(x^i,Y^i,f) + \frac{\|w\|^2}{n} + \Big(\frac{s^2\|w\|^2\ln(rn/\|w\|^2) + \ln\frac{n}{\sigma}}{2(n-1)}\Big)^{1/2} \quad (2)$$

for $\ell(x^i,Y^i,f) := \max_{y\in\mathcal{Y}}\{\lambda(Y^i,y)[\![v^i_y f(x^i,y) < 1]\!]\}$, where v^i is the binary indicator vector of Y^i.

Proof. The argument follows [1, Section 11.6], using the PAC-Bayesian bound

$$L(Q_w,D) \leq L(Q_w,S) + \sqrt{\frac{KL(Q_w,\pi) + \ln\frac{n}{\sigma}}{2(n-1)}}, \quad (3)$$

where π denotes a prior distribution on w, which we set as zero-mean Gaussian, $\pi(w) \propto \exp(-\frac{1}{2}\|w\|^2)$. We choose Q_w as a Gaussian centered at αw, $Q_w(\bar{w}) \propto \exp(-\frac{1}{2}\|\bar{w} - \alpha w\|^2)$. Then, the KL divergence between Q_w and π is just $\alpha^2\|w\|^2/2$. Analyzing the sample risk $L(Q_w, S)$ can be done for each training instance due to i.i.d. sampling. We denote by $\bar{Y}^i$ the predicted output for x^i with respect to $\bar{w}$. The proof is complete if we show

$$\mathbb{E}_{\bar{w}\sim Q_w}\Delta_{max}(Y^i,\bar{Y}^i) \leq \ell(x^i,Y^i,f_w) + \frac{\|w\|^2}{n}. \quad (4)$$

The claim follows then by inserting (4) and the expression for $KL(Q,P)$ into the PAC bound (3). □

To show inequality (4), we use

$$\begin{aligned}\mathbb{E}_{\bar{w}}\Delta_{max}(Y^i,\bar{Y}^i) &= \mathbb{E}_{\bar{w}}\max_y\{\lambda(Y^i,y)[\![v^i_y f(x^i,y;\bar{w}) \leq 0]\!]\} &(5)\\ &\leq \max_y\{\lambda(Y^i,y)[\![v^i_y f(x^i,y,w) < 1]\!]\}\\ &\quad + P_{\bar{w}}\{v^i_y f(x^i,y,w) \geq 1 \wedge v^i_y f(x^i,y,\bar{w}) \leq 0\}, &(6)\end{aligned}$$

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.

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:

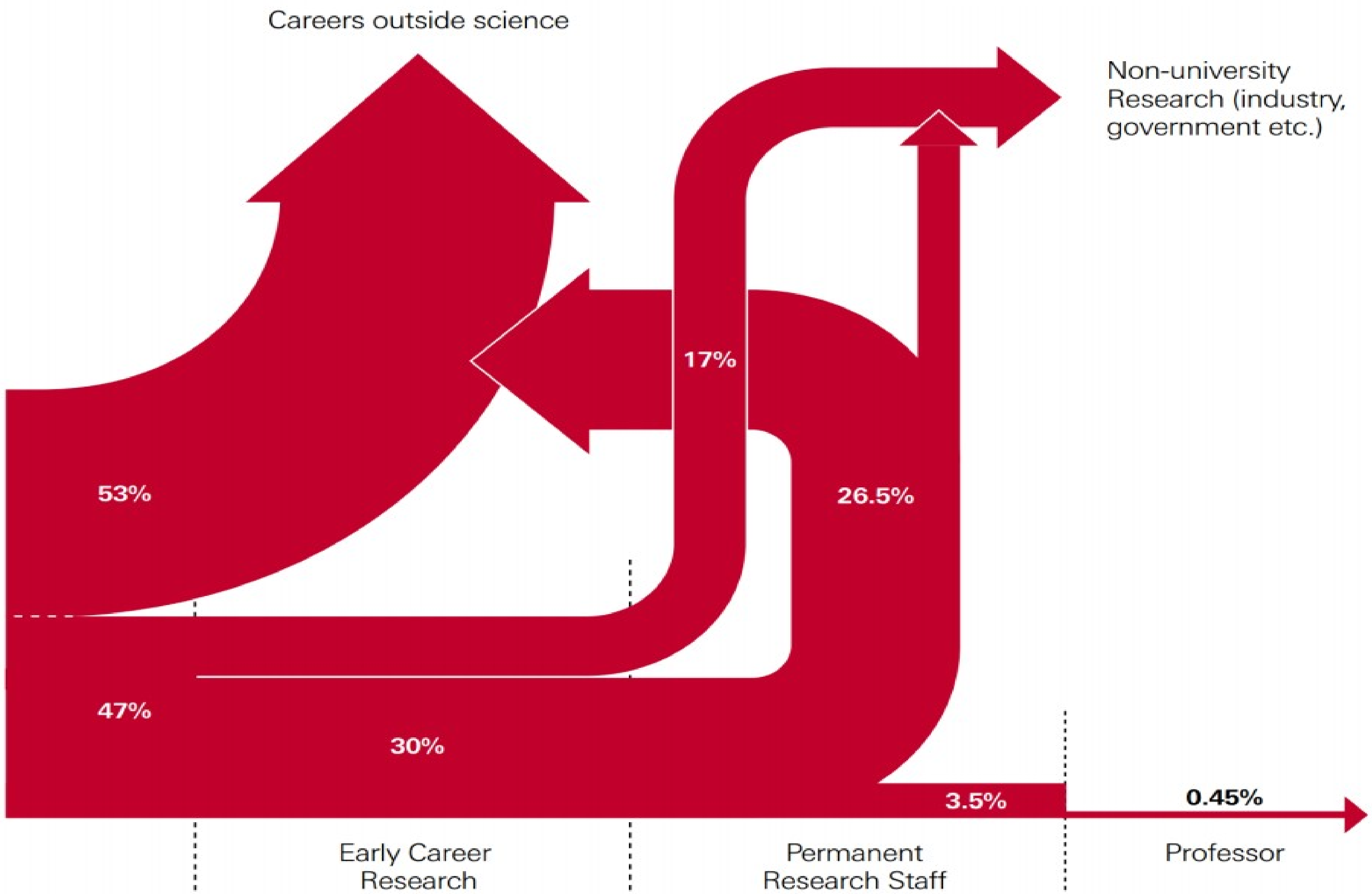
- Step 1: Obtain a Bachelor's and/or Master's Degree
- Step 2: Obtain a Doctorate/PhD
- Step 3: Work as a "Postdoc" for a few years
- Step 4: Become an Assistant Professor
- Step 5: Become a Tenured Professor

Outside of academia:

- Leave the process anywhere after Step 2

Beware: competition is fierce!

Careers of Scientists after the PhD



Source: The Scientific Century: securing our future prosperity, The Royal Society, London (2010)

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

Rank institutions in the world by publications from 2012 to 2022

All Areas [off | on]

AI [off | on]

- [ ] Artificial intelligence
- [x] Computer vision
- [x] Machine learning & data mining
- [ ] Natural language processing
- [ ] The Web & information retrieval

Systems [off | on]

- [ ] Computer architecture
- [ ] Computer networks
- [ ] Computer security
- [ ] Databases
- [ ] Design automation
- [x] Embedded & real-time systems
- [ ] High-performance computing
- [ ] Mobile computing
- [ ] Measurement & perf. analysis
- [ ] Operating systems
- [x] Programming languages
- [ ] Software engineering

Theory [off | on]

- [ ] Algorithms & complexity
- [x] Cryptography
- [x] Logic & verification

#	Institution	Count	Faculty
1	Carnegie Mellon University	25.3	89
2	Max Planck Society	20.2	26
3	ETH Zurich	19.4	37
4	Massachusetts Institute of Technology	15.4	71
5	Univ. of California - Berkeley	15.0	57
6	Stanford University	13.9	53
7	Cornell University	13.8	49
8	Univ. of Illinois at Urbana-Champaign	13.6	67
9	Univ. of California - San Diego	13.0	58
10	University of Pennsylvania	11.6	50
11	Tel Aviv University	11.3	31
11	University of Texas at Austin	11.3	27
13	University of Washington	10.4	46
14	EPFL	10.3	45
15	Technion	8.9	45
16	KAIST	8.8	54
17	IST Austria	8.4	11
17	Tsinghua University	8.4	68
19	University of Waterloo	8.1	24
19	University of Wisconsin - Madison	8.1	38

CSRankings: Computer Science Rankings

Rank institutions in Europe by publications from 2012 to 2022

All Areas [off | on]

AI [off | on]

- ▶ Artificial intelligence ☐
- ▶ Computer vision ☑
- ▶ Machine learning & data mining ☑
- ▶ Natural language processing ☐
- ▶ The Web & information retrieval ☐

Systems [off | on]

- ▶ Computer architecture ☐
- ▶ Computer networks ☐
- ▶ Computer security ☐
- ▶ Databases ☐
- ▶ Design automation ☐
- ▶ Embedded & real-time systems ☑
- ▶ High-performance computing ☐
- ▶ Mobile computing ☐
- ▶ Measurement & perf. analysis ☐
- ▶ Operating systems ☐
- ▶ Programming languages ☑
- ▶ Software engineering ☐

Theory [off | on]

- ▶ Algorithms & complexity ☐
- ▶ Cryptography ☑
- ▶ Logic & verification ☑

#	Institution	Count	Faculty
1	▶ Max Planck Society	20.2	26
2	▶ ETH Zurich	19.4	37
3	▶ Tel Aviv University	11.3	31
4	▶ EPFL	10.3	45
5	▶ Technion	8.9	45
6	▼ IST Austria	8.4	11

Faculty	# Pubs	Adj. #
Krishnendu Chatterjee LOGIC	49	14.5
Bernd Bickel GRAPHICS	33	6.6
Christoph H. Lampert VISION,ML	27	11.6
Thomas A. Henzinger LOGIC	25	7.0
Dan Alistarh ML	19	5.4
Christopher Wojtan GRAPHICS	18	6.2
Krzysztof Pietrzak CRYPTO	17	4.7
Vladimir Kolmogorov VISION,THEORY	9	3.2
Marco Mondelli ML	6	2.5
Gasper Tkacik ML	2	0.8
Calin C. Guet	1	0.2

#	Institution	Count	Faculty
7	▶ University of Edinburgh	6.7	39
8	▶ Ecole Normale Superieure	6.6	24
9	▶ TU Munich	6.0	23

ELLIS Society

News Events Research People Sponsorship About

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

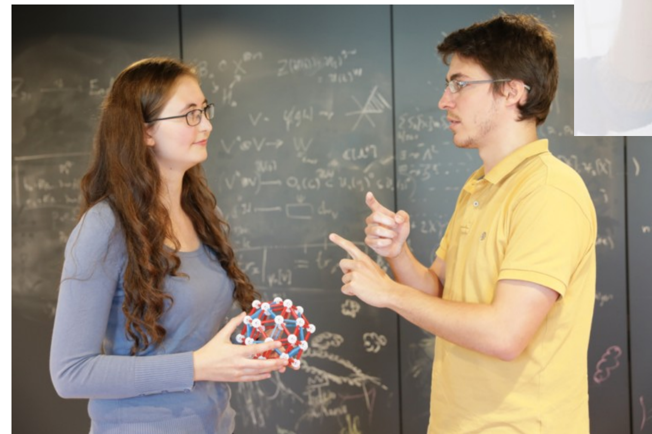
What to look for in a PhD program?

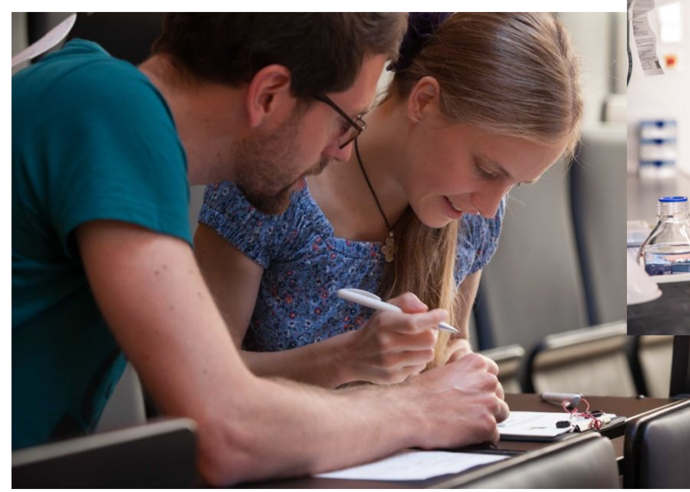
Running example:
the PhD Program at ISTA

Check the format

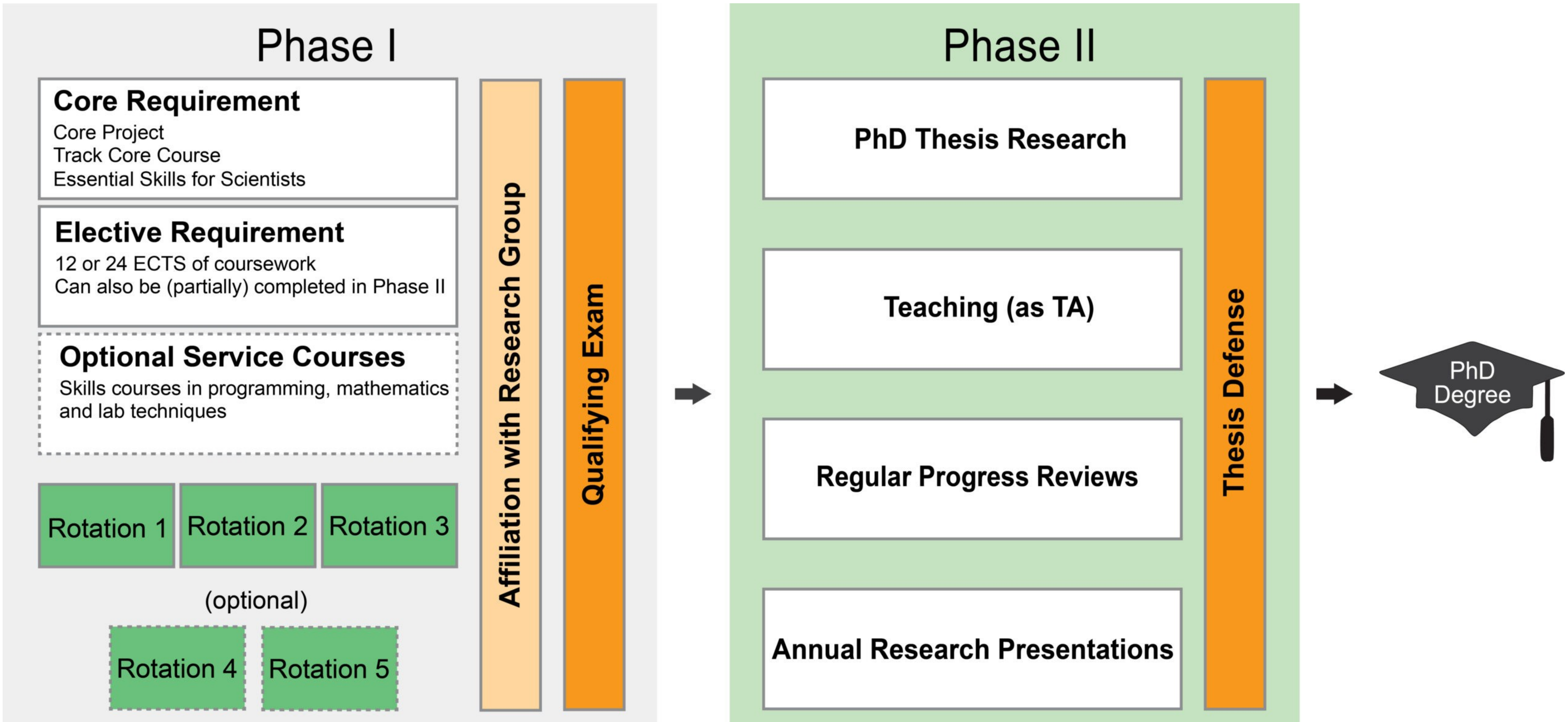
- Traditional European
  Master-Apprentice system
  - supervisor hires PhD student
  - usually requires Masters degree
  - few additional support structure

- (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



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

Photo: Ulf Liljankoski

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…

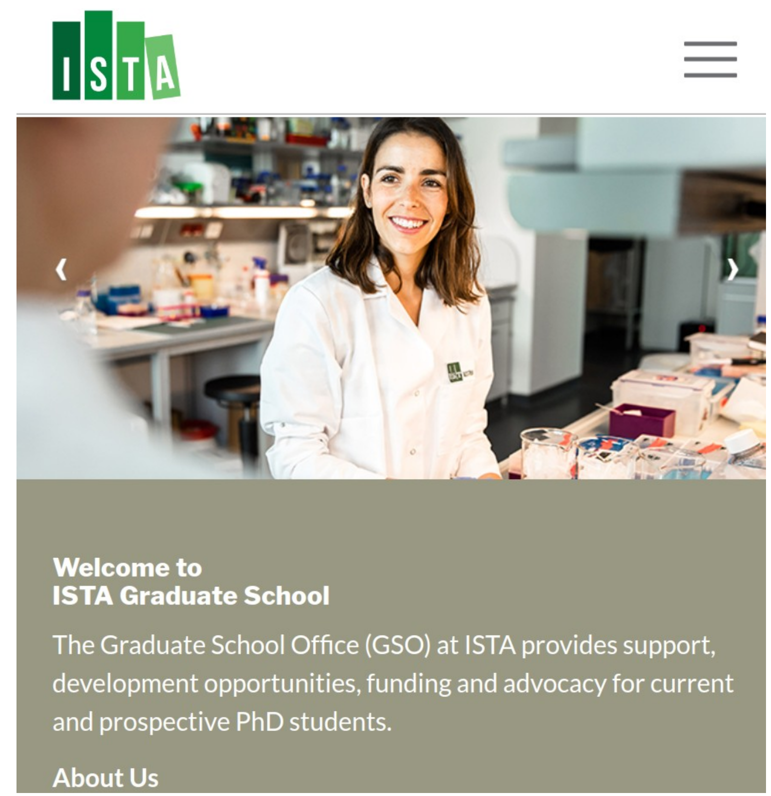
- ISTA PhD program: acceptance rate ~4% in 2021
  – over 4000 interested applicants
  – 2569 submitted applications
  – ~200 interviews (online and on-campus)
  – 106 offers
  – 67 accepted

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)

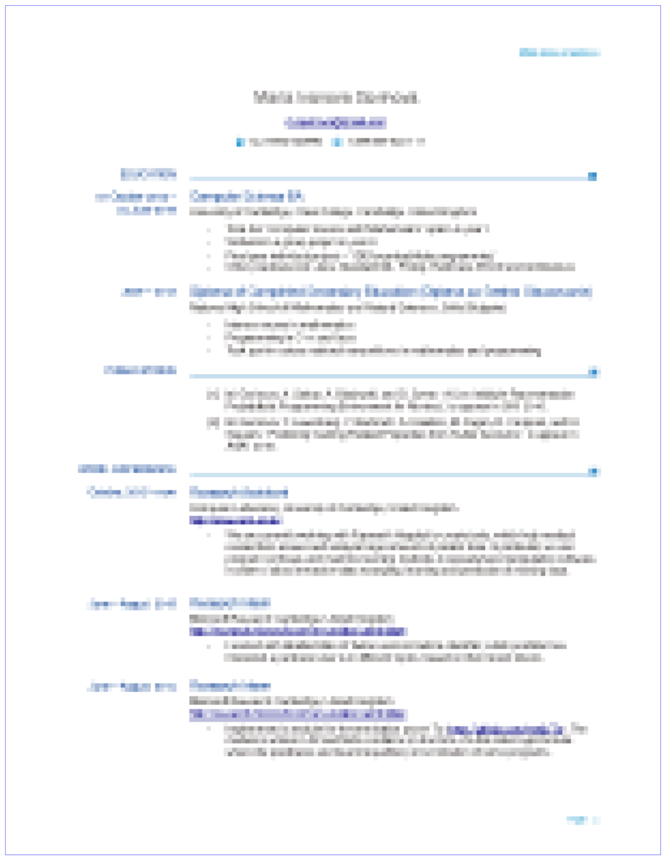
Getting into a strong PhD program

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.

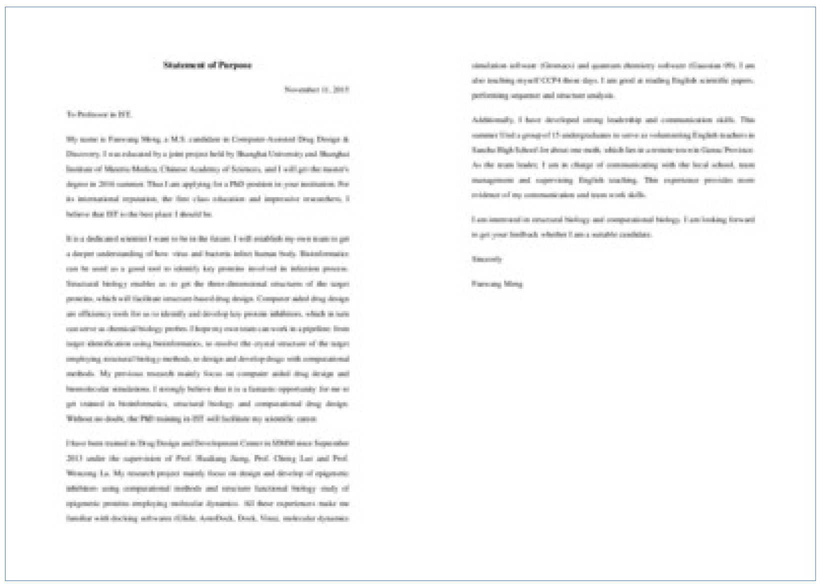
Getting into a strong PhD program

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


Statement of Purpose

looks good


STATEMENT FOR APPLYING FOR PHD

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

	ELLIS PhD Program
• 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?

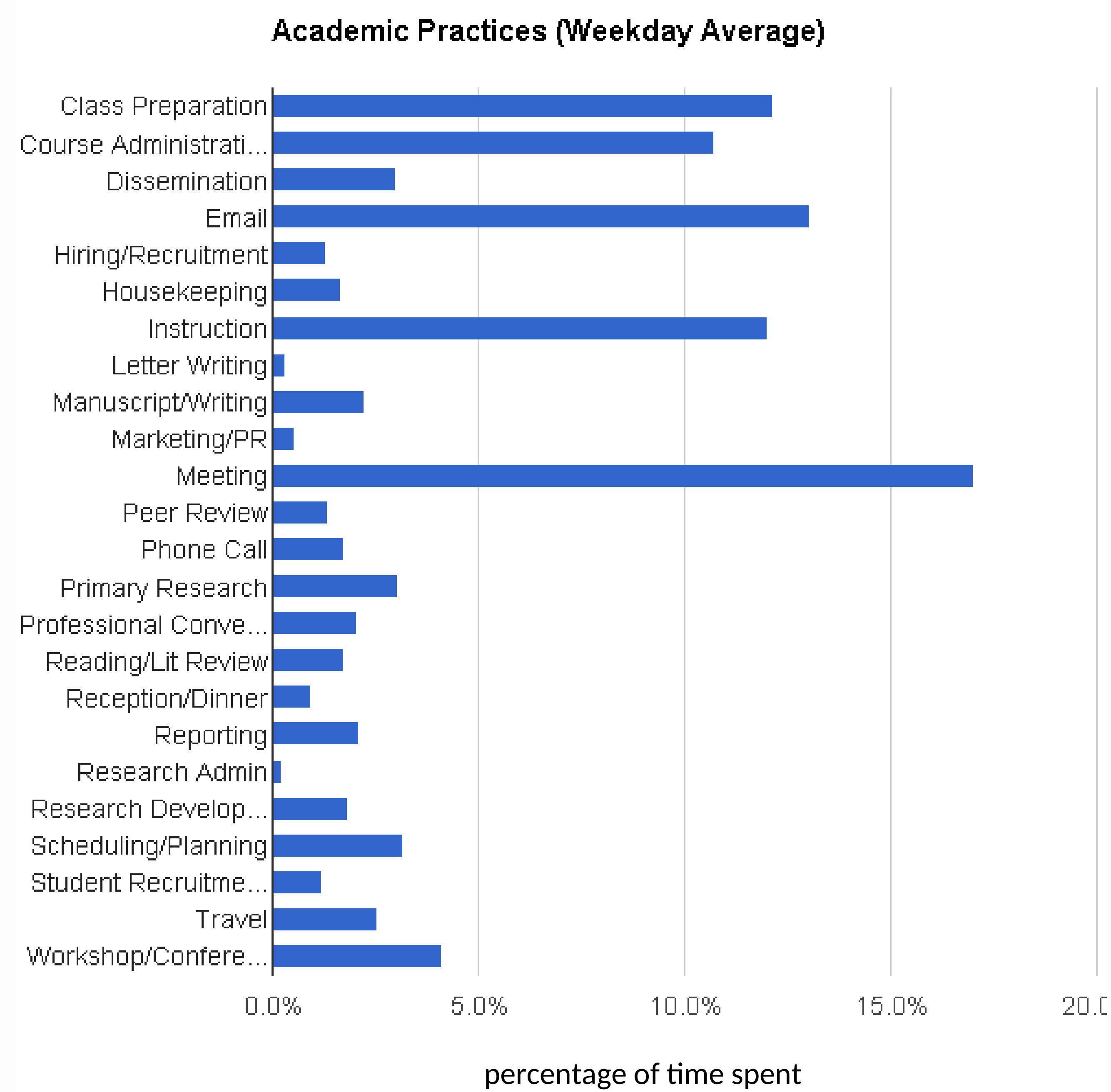
- summer schools, internships, vacation, volunteer work..

https://ellis.eu/phd-postdoc


QUESTIONS
ANSWERS

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

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

PhD student

Education

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

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

PhD student

Education

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

How do researchers spend their time?

Researcher in industry

Postdoc in academia

Administration

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

Research

- actual real research
- publications:
  - writing
  - reading
  - reviewing

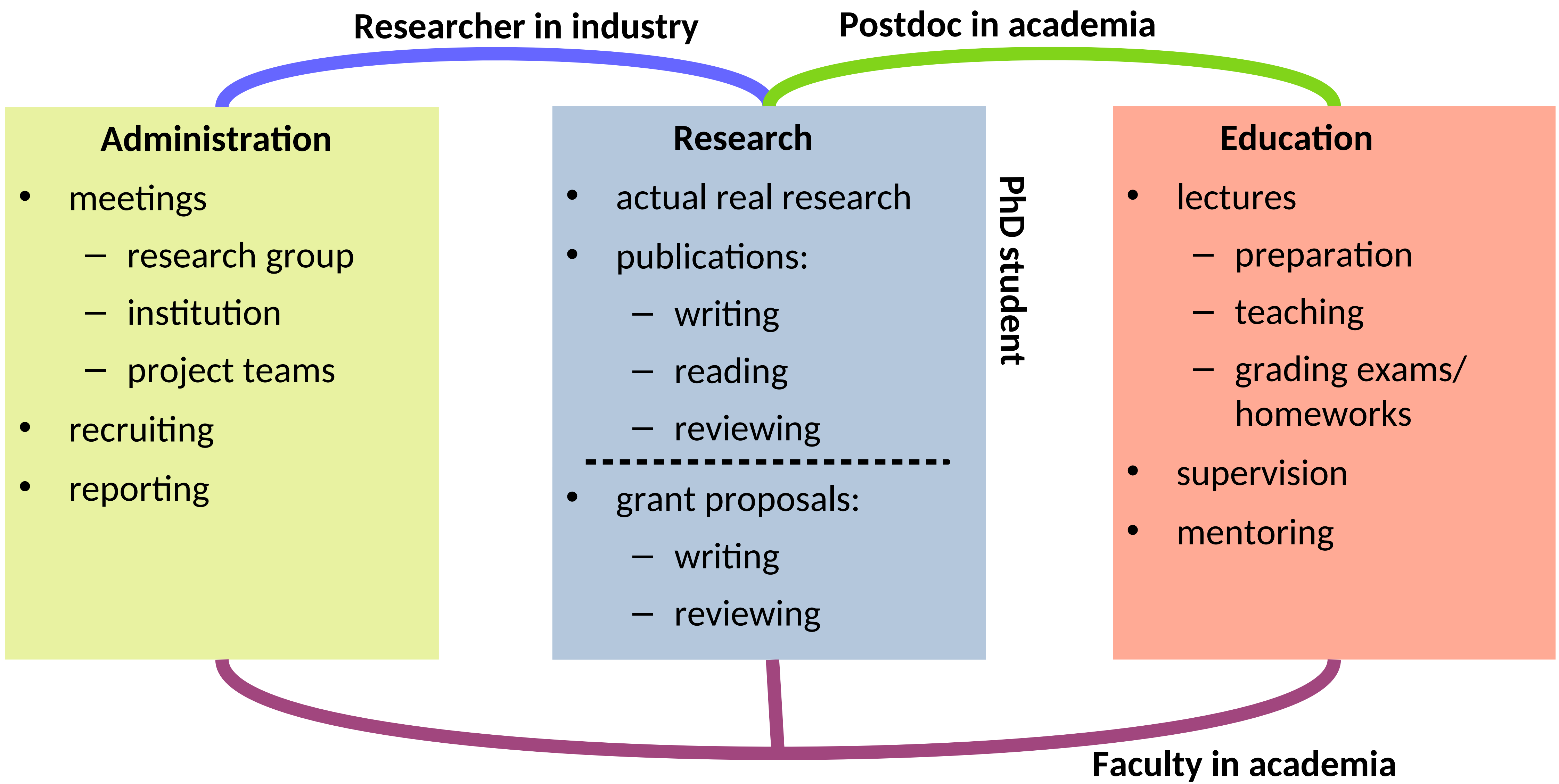
- grant proposals:
  - writing
  - reviewing

PhD student

Education

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

How do researchers spend their time?



How do researchers spend their time?

Researcher in industry

Postdoc in academia

Administration

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

Research

- actual real research
- publications:
  - writing
  - reading
  - reviewing

- grant proposals:
  - writing
  - reviewing

PhD student

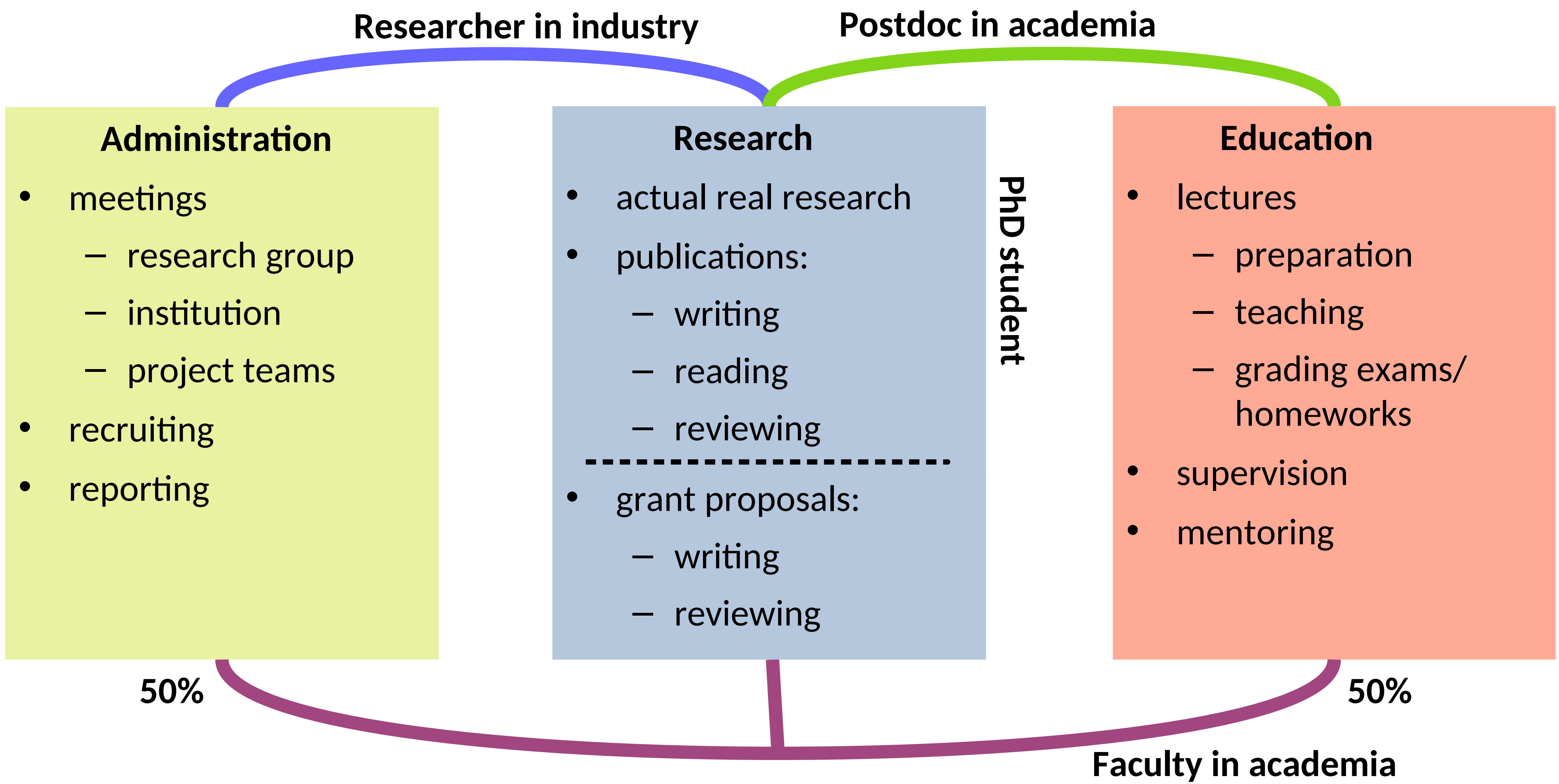
Education

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

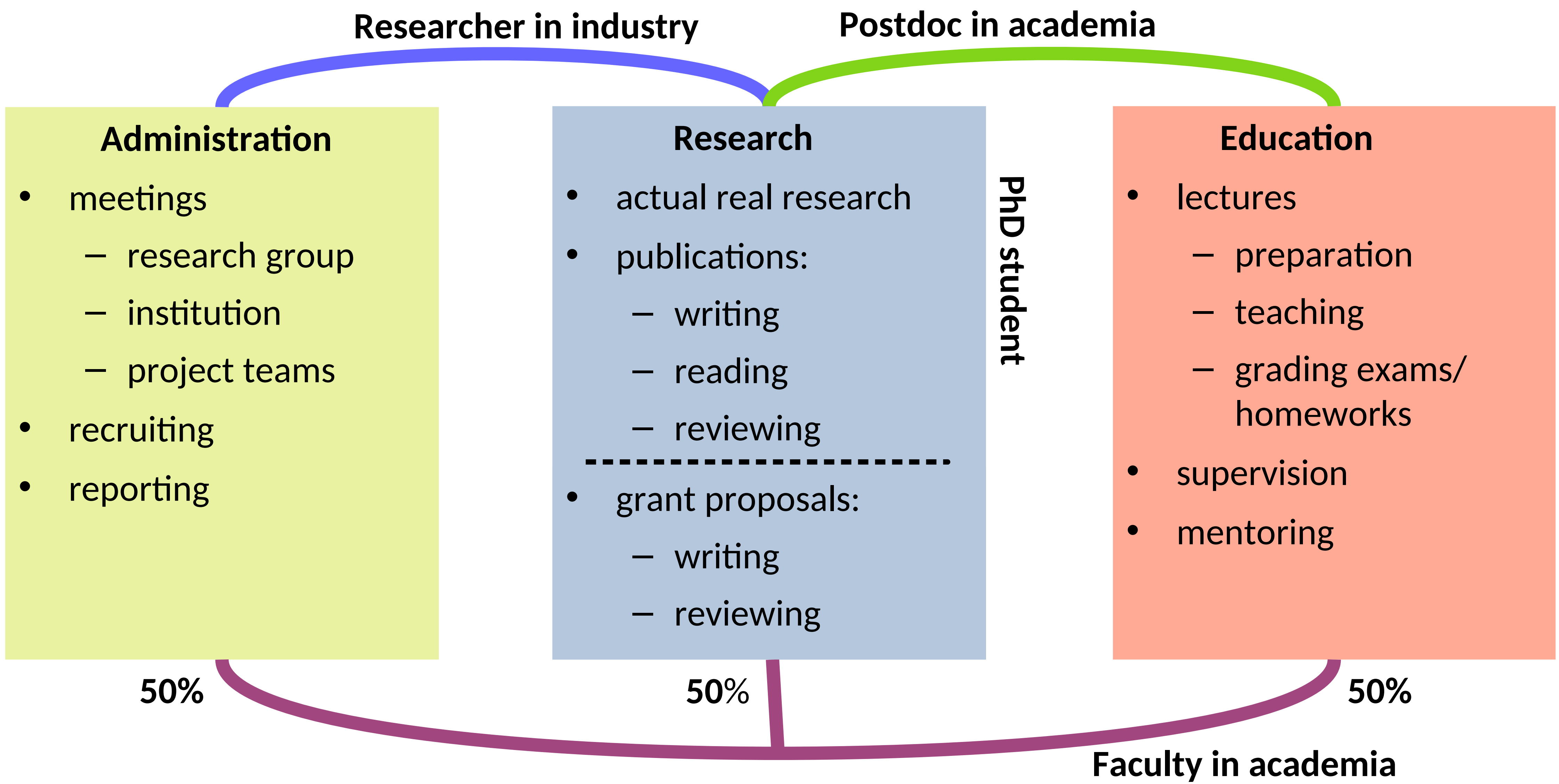
50%

Faculty in academia

How do researchers spend their time?

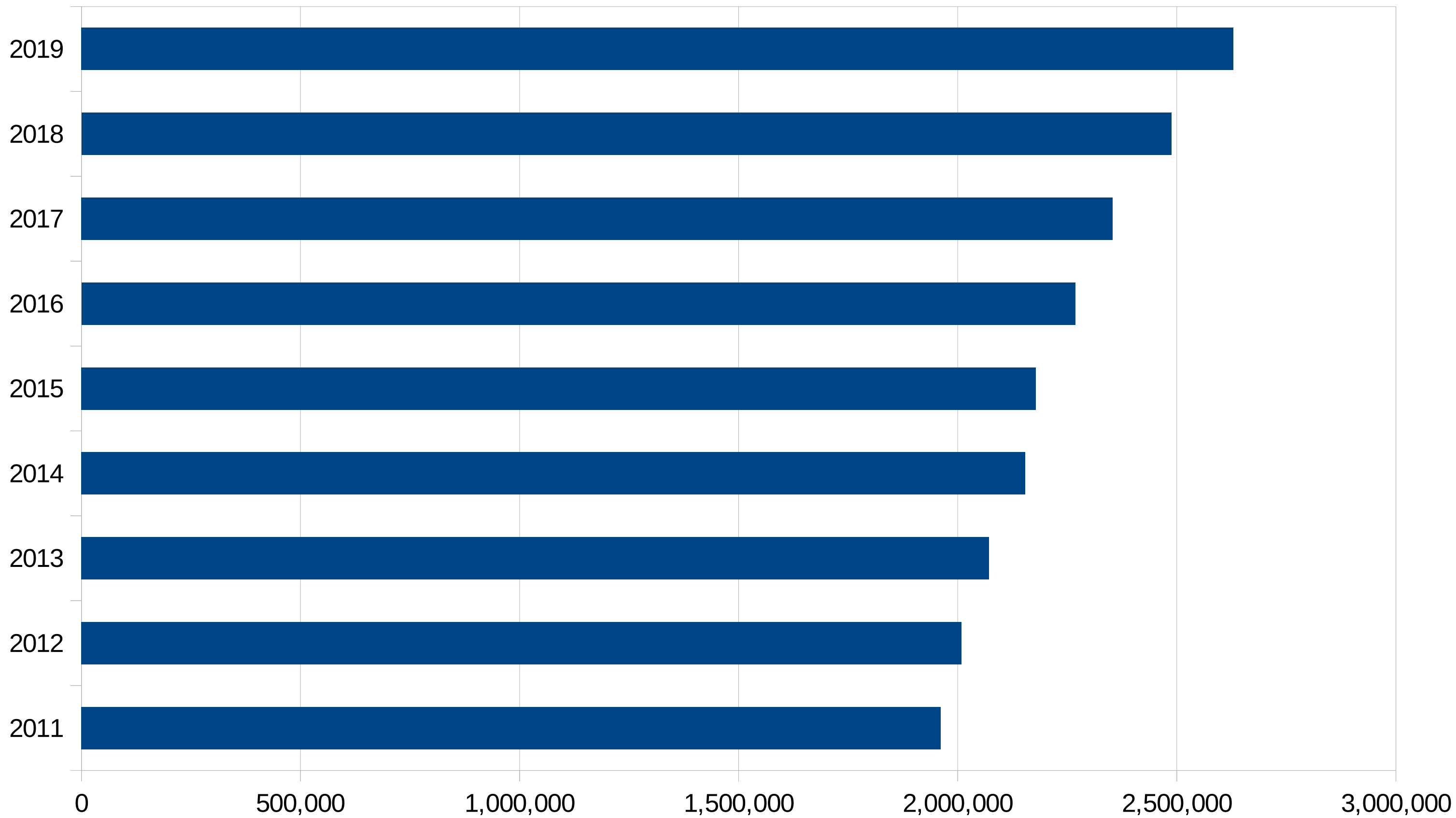


How do researchers spend their time?



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

Source: UNESCO (2021) UNESCO Science Report: the Race Against Time for Smarter Development. S. Schneegans, T. Straza and J. Lewis (eds). UNESCO Publishing: Paris.

Behind the scenes: Publish or Perish

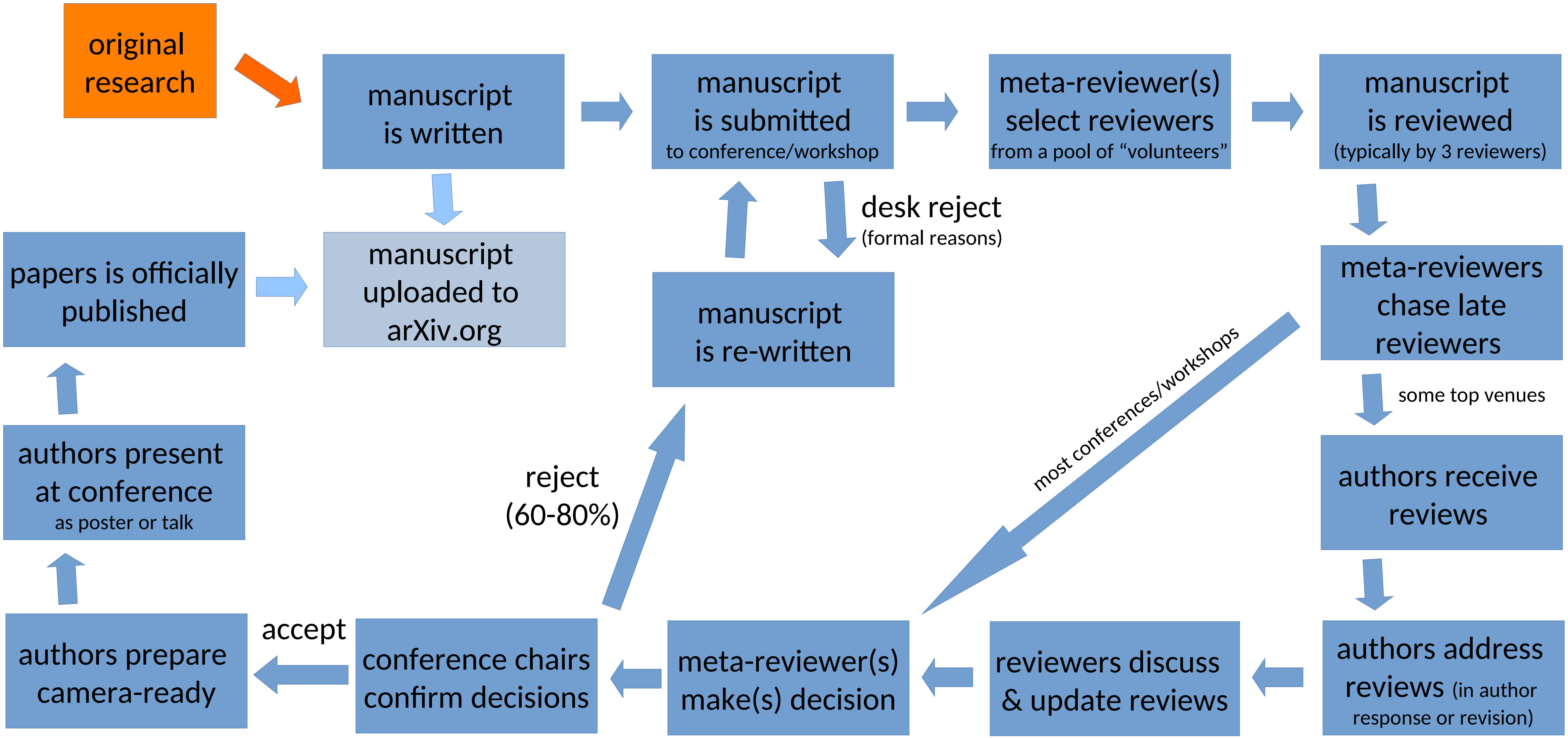
In a large research area, such as machine learning, it's impossible to stay up-to-date with all works!

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

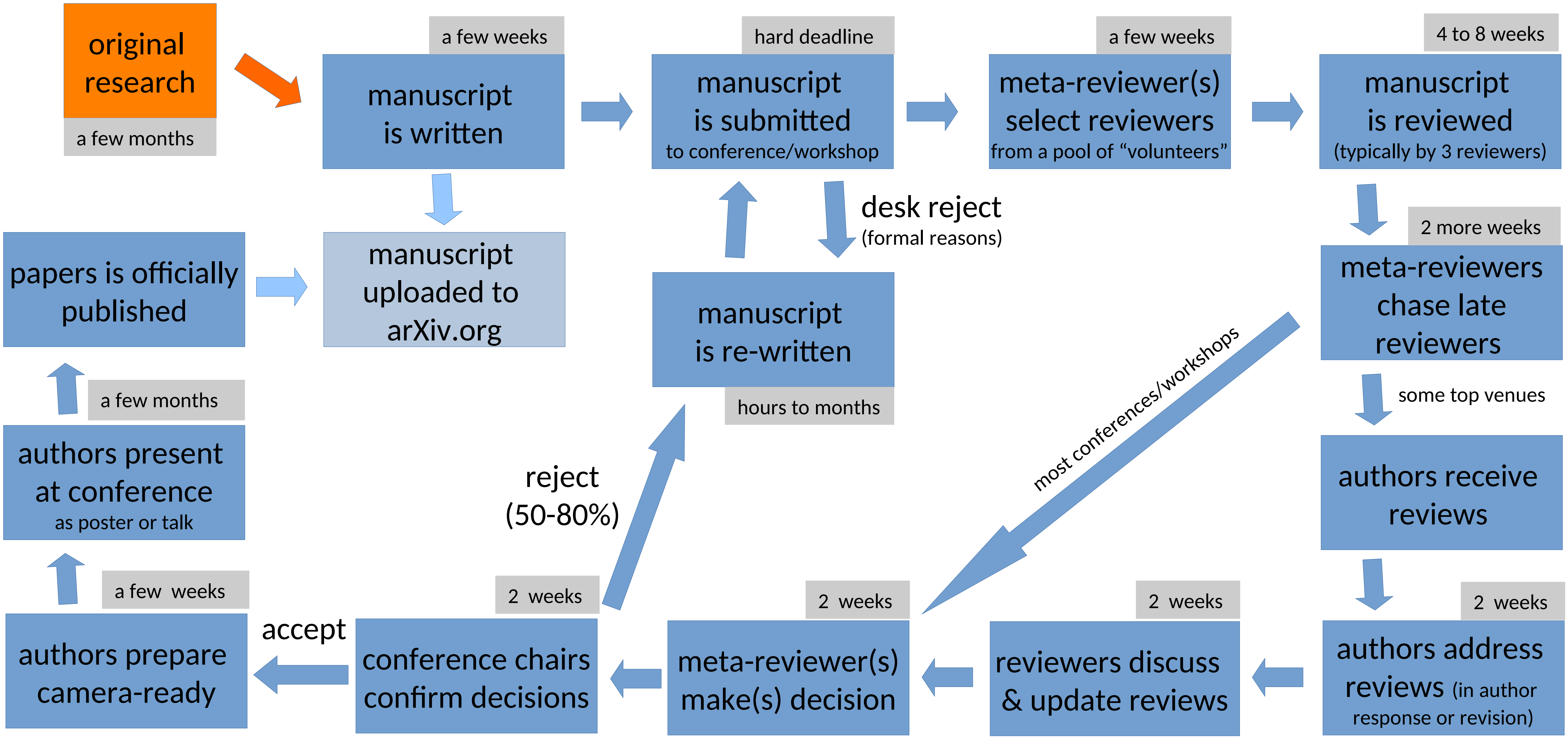
	Peer Review	Name Recognition	Aggregators	Social Media
what to read?	what is published at top conferences or journals	publications from top research labs	automatic digests, e.g. Google Scholar, based on keywords or citations	what shows up on Twitter or Youtube
how much?	a few thousand papers every couple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

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…

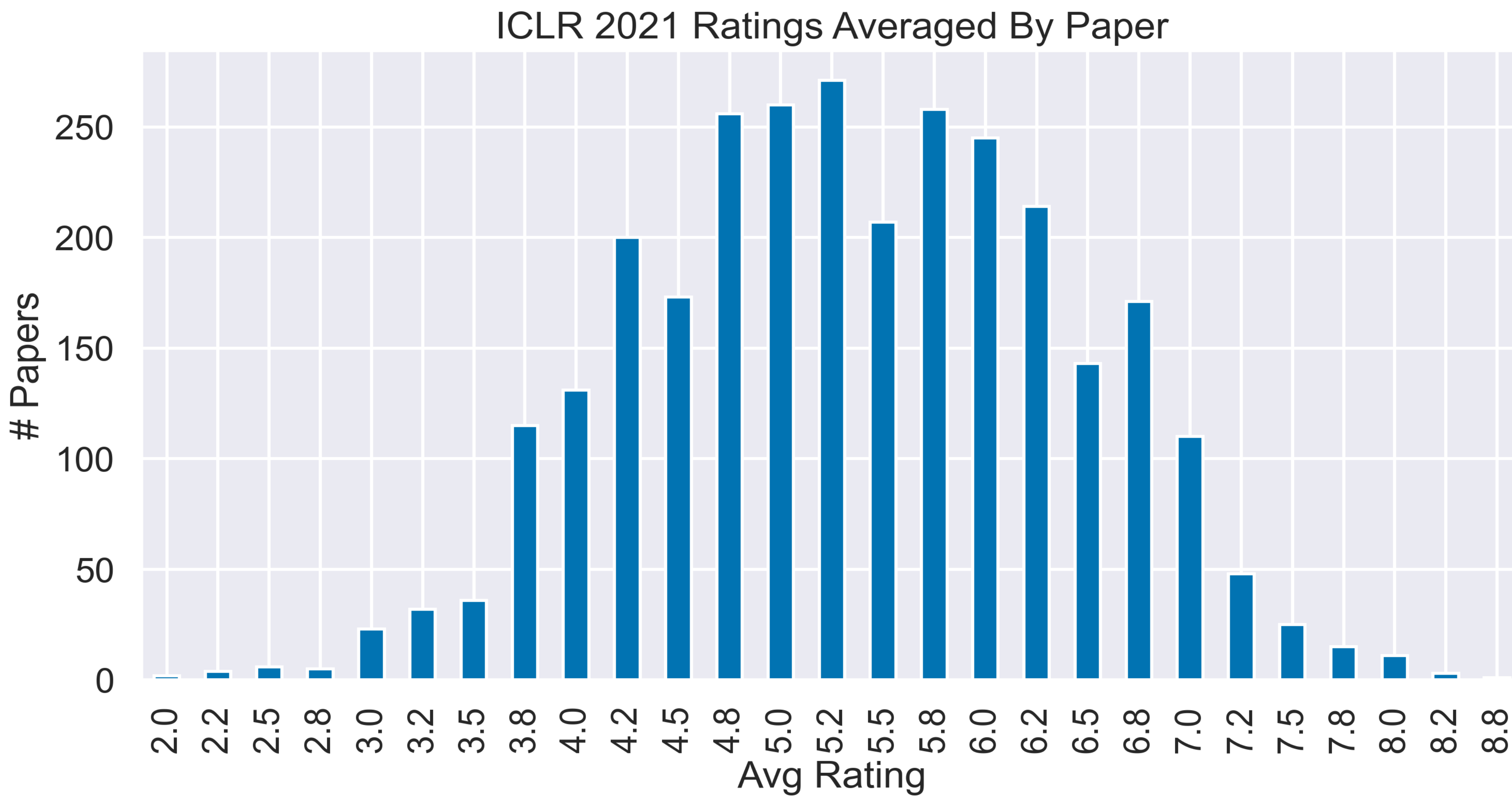
The long path of a (machine learning) publication



The long path of a (machine learning) publication

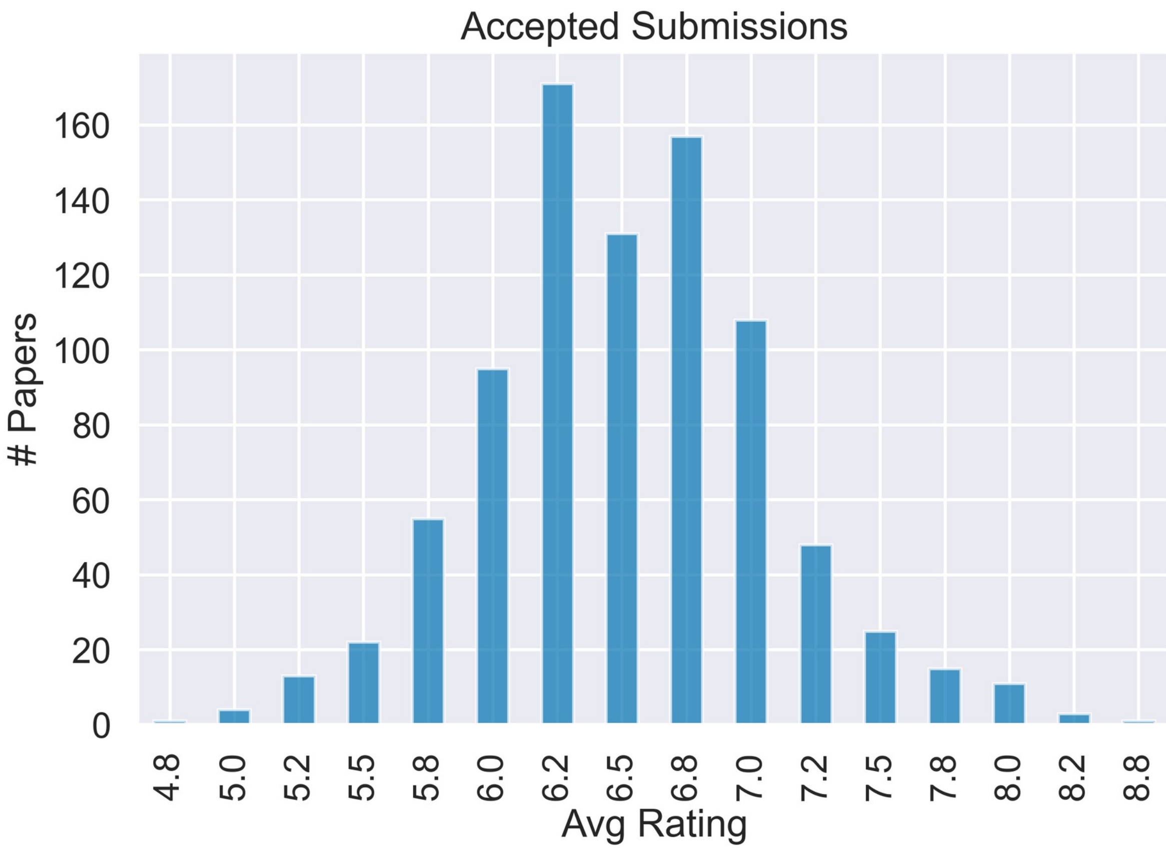


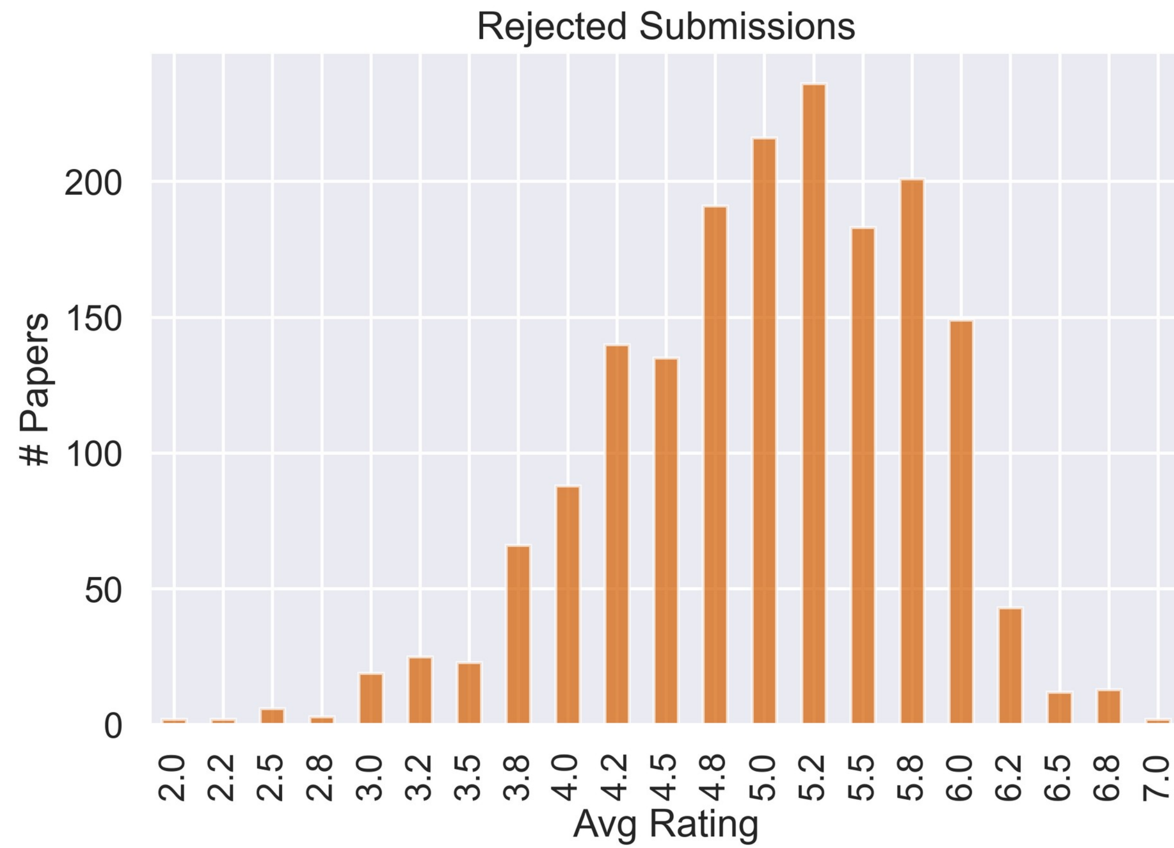
Behind the scenes: Publish or Perish



Source: https://github.com/evanzd/ICLR2021-OpenReviewData

Behind the scenes: Publish or Perish





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

1st \ 2nd	**Accept**	**Reject**
Accept	99	107
Reject	96	462

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)

Source: Congressional Research Service, https://crsreports.congress.gov, R44283

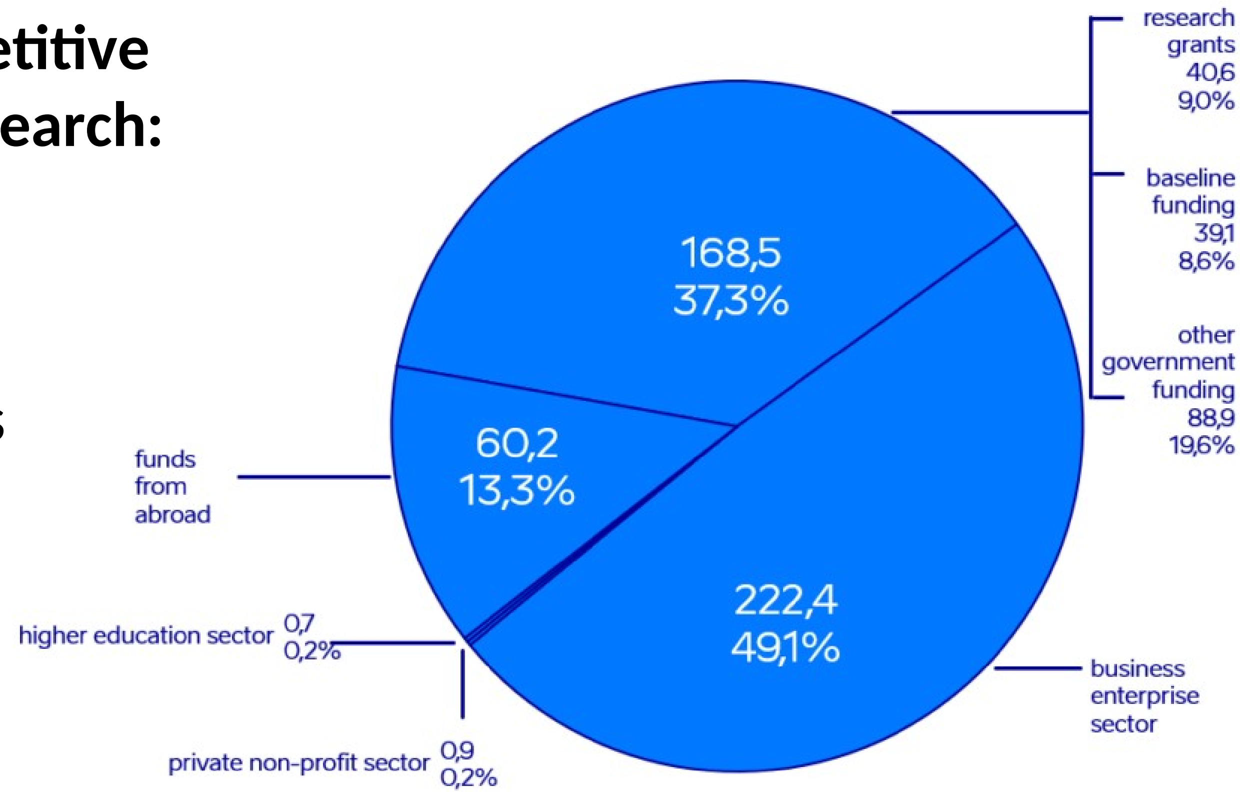
Research Funding

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:

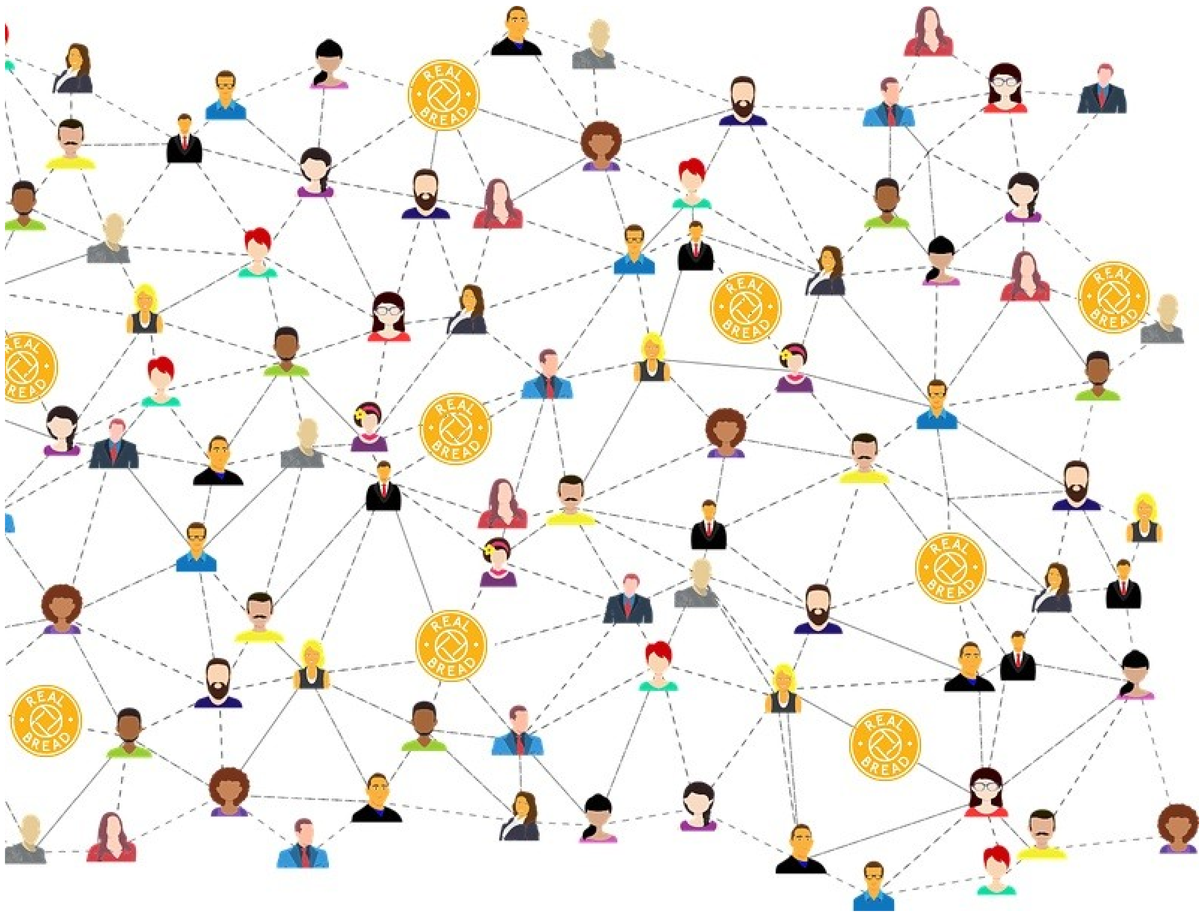
- time-sheets: who worked on which parts of the project for how long
- intermediate reports about milestones and deliverables
- intermediate review meetings with external experts

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!


QUESTIONS
ANSWERS