Report of the 2021 Area Evaluation Panel for Data and Computational Sciences

Martin Abadi, Google AI
Kavita Bala, Cornell
Peter Dayan, MPI- Biological Cybernetics
Gene Myers, MPI- Cell Biology and Genetics
Sylvia Ratnasamy, UC Berkeley
Stuart Russell, UC Berkeley (Chair)

In a very short time, IST has succeeded in establishing high-quality programs in data and computational sciences, with many world-class faculty. Faculty morale is excellent. The panel agrees with IST's general plans to broaden the computer science effort beyond its current, mostly theoretical, footprint. Specifically, we recommend at least 12-15 CS hires to create a broad, world-class department.

Like many institutions, IST will need to think through the implications of the explosive growth in computing's economic, social, and intellectual impacts. For data science, it will need to decide if the label corresponds to a lasting disciplinary core involving data, modeling, and inference, and, if so, will need to hire 6-10 data scientists (primarily in statistics and AI/ML) to build up that core.

1. What is the international standing of data and computational sciences at IST Austria, also relative to other centers in Austria?
    a. What are its strengths and weaknesses?
    b. How could the weaknesses be addressed and what would be appropriate strategic goals for data and computational sciences at IST Austria?
    c. Is there a minimum size and the necessity to cover certain fields in order to compete for a spot among the top 10 data and/or computational sciences groups in Europe?

We feel it necessary to clarify terminology regarding disciplines, as "data and computational sciences," while descriptive, is not yet a recognized disciplinary label. Instead, we will talk about computer science and data science, the former a well-recognized discipline and the latter an emerging field comprised of machine learning, parts of database systems and theory, statistics, and domain scientists in various data-centric disciplines (biology, high-energy physics, political science, climate science, and so on) who typically contribute to the development of computational and statistical methods as well as their use.

The panel finds it difficult to place IST relative to other leading centers for data science and computer science, because we evaluated only a portion of faculty members in data science and computer science and in both areas (especially data science, where statistics is usually regarded as central) IST's coverage is still quite partial even if we include non-evaluated faculty.

Despite these caveats, we are extremely impressed that in just over a decade, IST has attracted a remarkably strong collection of computer scientists. Taken as a whole, this is a world-class group, with particular strength in formal methods, two stellar early-career graphics researchers who also contribute to scientific computing, and a trio of high-quality, theoretically inclined researchers who focus on machine learning and distributed algorithms. The two young domain scientists we evaluated (in biomedicine and neuroscience) are also very strong.

Also notable are the excellent morale and esprit de corps of both faculty and students. IST is a well-designed and well-run institution. The "growing pains" noted by some faculty members are to be expected as IST's increasing size necessitates the creation of internal organizational structures.

IST's primary weakness in computer science is simply a consequence of its institutional youth: the lack of faculty in experimental core CS (including systems) and in applied areas other than graphics. Due to the partial nature of the evaluation we are uncertain of IST's strength in other important theoretical areas such as algorithms, complexity, cryptography, and quantum computation.

A related consequence of institutional youth is the difficulty in attracting top graduate students and postdocs, who often make decisions based on institutional reputation. Although some of the evaluated faculty were happy with the quality of students, others were frustrated by losing their best prospects to other institutions. Several of the students we spoke to had heard of IST only through personal contacts or lectures by IST faculty. More effort should be expended (it is already a priority clearly) to build strong pipelines to various institutions across Europe and internationally to attract the best student talent possible. The internship programs, both IST-wide and lab- and group-specific, should be expanded and outreach strengthened. Cultivating an elite image (enhanced by guaranteed funding offers, plenty of faculty access, and opportunities to teach at the graduate rather than undergraduate level) may be helpful in attracting more applications and raising offer acceptance rates.

The present landscape of computing in the world is very different from 40 years ago, when isolated computers existed to process batch jobs entered on punched cards in data processing centers; IST's hiring needs to reflect this. Now, billions of people carry in their pockets more computing power than most nations could buy in 1981, and information flows have grown more than a billionfold. We note that the hiring of Lefteris Kokoris-Kogias is a welcome step towards reflecting all these changes.

Whereas an all-male faculty might have been typical in 1981, it is not really acceptable in 2021. IST's report acknowledges this issue and details significant efforts to address it; we suggest some possible approaches later in this report.

We estimate that hiring 12-15 additional faculty comparable in quality to the existing CS faculty would put IST clearly among the top CS departments in Europe. It would also be enough to support a comprehensive graduate education program and a graduate student body large enough to benefit from it. IST's report includes suggestions for areas, some (security, programming languages, databases, networks) more plausible than others (computational geometry, combinatorics). In addition to hiring new faculty, it will be important to adjust the PhD recruiting process, which currently seems to attract a high number of mathematicians from Eastern Europe and Russia who may not be ideal candidates for non-theoretical areas.

In Austria, TU Wien is the strongest CS department. It has roughly 70 faculty members with fairly broad coverage (perhaps lacking in modern AI and machine learning). In its current areas of coverage, IST ranks higher; if and when IST's coverage of core areas of CS becomes more complete, it will undoubtedly be seen as a stronger institution than TU Wien for computer science.

In data science, the situation is somewhat different, partly because of the lack of a generally recognized disciplinary structure. Different institutions around the world have treated data science in different ways. Some have ignored it completely, perhaps feeling that the use of data and models in the natural and social sciences is hardly a new phenomenon that merits big organizational changes; some have augmented the applied and computational side of their statistics departments; some have strengthened ties between statistics and CS; and some have created entirely new programs, colleges, or schools, trying to bring together CS, statistics, and domain scientists. If there is a disciplinary core to data science, it lies in the connection between data and models, i.e., in statistical inference and machine learning. At IST, engagement between machine learning faculty and domain scientists is not yet systematized, and the core areas of probability and statistics are not yet developed. The panel did not hear a coherent strategy for organizing or developing data science. We welcome the suggestion to add groups in "high-dimensional statistics, as well as practical aspects of machine learning"; but this should be seen as essential as part of the core for data science, not for the purpose of developing a hub in theoretical deep learning. Deep learning is just one tool, not an end in itself.

If IST decides to develop its strength and presence in data science, it will probably need to hire 6-10 faculty in statistics and applied machine learning (some of whom may overlap with the CS hires). The statisticians would probably be in areas such as computational statistics, inference, or large-scale Bayesian modeling rather than more traditional areas such as significance testing.

It may seem odd to remark on this, given that IST's President is a computer scientist, but the panel feels the need to emphasize that CS is a discipline of intrinsic value, not a service discipline for the "real" sciences. CS emerged from Turing's discovery of new kinds of mathematical objects - machines and programs - that turned out to be the most powerful mathematical objects ever found, even more so than numbers. CS systems research, in collaboration with solid-state physics and electrical engineering, turned mathematics into reality. In the last few decades CS has created 8 of the 10 largest companies in the world and dramatically changed human lives; its future impact through AI may be far greater. In many

universities around the world, students have recognized this, flocking to courses and degree programs, while university administrators struggle to adapt legacy organizational structures and reallocate resources to reflect these developments. IST is in the enviable position of starting from a blank slate and has the opportunity to develop programs and structures that achieve a proper balance and enhance the opportunities for collaboration and cross-fertilization.

2. In the near future, specifically with regard to new hires:
    a. Shall IST Austria focus on its existing strengths or expand into new subfields of data and computational sciences?
    b. What are subfields of data and/or computational sciences and adjacent directions that would fit particularly well to the research portfolio of IST Austria?
    c. Are there any subfields of data and computational sciences that you recommend not to pursue?

The panel felt strongly that it was important for IST to branch out from its more theoretically focused work. In addition to making IST a more complete and attractive institution for computer science research and facilitating connections to and development of industry partners, this would both complement and strengthen the existing theoretical research portfolio by ensuring the relevance of theoretical results and motivating new theoretical investigations to support the solution of real problems.

This branching out could mean ***experimental*** (or hybrid theoretical/experimental) research areas in core, application-independent computer science, including security, distributed systems, operating systems, networking, databases, and HCI; or it could mean research in ***applications*** of computing in areas such as applied machine learning, computer vision, robotics, computational systems biology, e-commerce, social media, and possibly computational precision health (if there is a suitable medical-school partner in Vienna). Note that graphics, an existing strength, is an application area.

Specifically, we already heard there was an interest in systems and robotics hiring. For both areas, a cluster hire might be best. For systems, it would be good to keep in mind potential synergies with the formal methods area. Systems is a broad area and certain sub-areas would benefit from adjacency to existing IST faculty: e.g., security and programming languages would complement Henzinger/Chatterjee while networking/OS would augment Alistarh/Kokoris allowing a nexus to be built around provably robust systems (an area that will face increasing demand in practice). For robotics, the absence of faculty in control theory and mechanical engineering may make it challenging to build a viable group. On the other hand, this might be a good place to start building up the T in IST.

We heard a desire to set up as a hub of principled deep learning, but the panel recommends aiming for a broader approach looking beyond deep learning, with a focus defined by broad problem areas (e.g., perception, decision making) rather than methods.

A concern that was raised was whether IST is too risk-averse in its hiring. That seems likely to be the case. People tend to hire other faculty with similar tastes. This makes breaking out into a new type of area (applied or experimental instead of theoretical) difficult. Various suggestions were made to try to expand the range of area covered:

- Setting up a distinguished speaker series in each plausible new area to educate the IST faculty as to what “good” research in this area looks like. This is important because there can be non-trivial cultural differences between theoretical and experimental/applied scientists in terms of the nature of innovation and measures of success.
- Asking several area experts (spanning geographies and seniority levels) to be actively involved in the recruiting process to improve the odds of a successful outcome.
- Hiring in a cluster to create critical mass in an area, thus ensuring that multiple offers get accepted and decreasing the risk for a person who has to decide to be the “first” person in the area to come to IST. The appropriate cluster size will vary depending on the area and should be carefully considered.

The panel did not recommend hiring in areas such as experimental computer architecture/hardware that require substantial investment and struggle to compete with industry research or even to be relevant. Although quantum computing is an exciting area, creating a successful group might require coordinated hiring with the Physics faculty, which may be tricky to pull off.

Beyond the above discussion about hiring to branch out, other topics should be addressed in the hiring process.

In terms of salaries and competitiveness, there was agreement that Swiss salaries were higher, and there was no expectation from the administration that the salaries would be matched. However, some junior faculty mentioned that while the cost of living was lower in Austria, the taxation system meant that the salaries were perhaps not competitive with respect to Swiss salaries even given the cost-of-living considerations. We recommend an analysis of the salaries to make sure that they are indeed reasonably competitive and correctly handle hidden costs like taxation. It might be useful to define a formal comparison group and regularly update the comparative data.

The hiring process seemed quite onerous with significant obstacles to moving quickly on offers because an extra round of letters from external letter writers is solicited even for junior faculty. This must be addressed to make the hiring process more nimble.

Building out the technology park through more industry, support for startups, etc. could result in a more vibrant ecosystem that could help address two-body situations. Possibly IST has already carried out a systematic study of potential Austrian or foreign industry partners; if not, this would be a useful exercise.

3. For each one of the data and computational sciences professors,
    a. What are their individual strengths and weaknesses as a scientist?
    b. How does the performance of each data and computational sciences professor compare with what would be expected from a faculty member in the same field and of the same seniority at a top 10 data and/or computational sciences department in the US?
    c. Do the assistant professors receive enough support from the Institute to reach their full potential as scientists?

Evaluations of individual faculty members follow. In response to 3(c), we note that some junior faculty members did not know they were supposed to have both internal and external mentors. This kind of process bug is not unexpected as IST's academic personnel procedures evolve and the area chairs' roles and responsibilities solidify. Other than this, the junior faculty feel very well supported, have more than enough funding, and find the atmosphere at IST to be open and encouraging.

[Note from Academic Affairs about pages 7-16: The sections on individual professors of the review report are treated in the same way as confidential review letters. Hence they are not included in this version of the report.]

4. How do data and computational sciences contribute to the larger goal of IST Austria to become an internationally recognized multi-disciplinary center of scientific excellence? How does the Institute, its structures, and its culture support data and computational sciences as disciplines, and how could the support be improved?

   What do data and computational sciences contribute to the Institute, its culture, and to other disciplines at the Institute, and how could the contributions be improved?

   How do you assess the doctoral education at IST Austria, in particular the multidisciplinary PhD program, with regard to data and computational sciences, and how could it be improved?

The mission of IST Austria is to perform research and doctoral education on a world-class level, and the investments in data and computational sciences have contributed substantially to this mission. These areas have been prominent in IST since the start; they were key to putting IST on the map. Work on these areas continues to receive clear recognition, with various awards and grants (e.g., prestigious ERC grants). Since data and computing underlie so much of the modern world, and since there is so much growth and change in these areas, the relevance of data and computational sciences is evident. At the same time, it would be insufficient, and detrimental to IST's mission, to think of data and computational sciences exclusively in terms of their applications in other science and technology fields and elsewhere. Data and computational sciences include areas with their own rich and important scientific agendas, including decidability, universality, correctness, complexity, approximability, cryptography, learnability, and inference; IST is already established as a center of scientific excellence in some of those areas, and the further growth of IST should reinforce that excellence and expand it into other areas. (Some of those areas should include significant experimental aspects, as we discussed in the answer to Question 2: "experimental" does not necessarily mean "applied" in the sense of "focused on applications".)

IST has provided excellent support for data and computational sciences in its first dozen years. The creation of divisions within IST should enable scaling. The divisions—and, over time, more structure or specialized faculty committees—may enable IST to recognize the specificities of each area and to support all areas appropriately. (As one small example: acceptance rates for faculty candidates in Computer Science are lower than in many other fields represented at IST, so perhaps having relatively more interviews in Computer Science would be reasonable, and support the effort to improve gender representation.) IST is paying appropriate attention to ensuring that the new structure is not an obstacle to collaborations, including important collaborations between data and computational sciences and other fields, e.g., biology, which are essential to its multidisciplinary mission.

The panel's exposure to the PhD education program was limited, but it is evident that the PhD students we met feel well-supported and enjoy the collegial atmosphere at IST. For most of the research groups we evaluated, the quality of thesis research and the post-completion job placements suggest that the program works well even though the graduate course offerings are

necessarily limited. Although it is no longer possible for a PhD student to have an in-depth understanding of every major area of CS, the panel felt that the current course offerings could be strengthened to ensure that students have exposure to the major modes of thinking in CS research, including theoretical, experimental, and applied research.

The general IST-wide onboarding course is a fairly unique feature; unfortunately the panel did not probe students on how they perceived this course and how valuable it was for them. It is also unclear to us how well the course succeeds in conveying to students in physics and biology the beauty and depth of computer science, as distinct from its utility in processing their data.

Questions Specific to Data and Computational Sciences

1. What measures do you suggest to attract women faculty?

Hiring women has been a stated priority for IST Austria since 2012, and it seems a more serious effort has been made in the past few years with the institution of a specialized committee to drum up applications from women and manage the process. The hiring committee reported their disappointment that despite significant efforts they were unable to convert offers to women joining IST, reporting that they lost candidates to MIT, CMU, but also to perhaps less competitive schools such as BU. The reasons are varied including lack of viable two-body solutions, partly because of the lack of a vibrant technology ecosystem for spouses to find jobs. They do note that faculty who come to IST like the quality of life and hope that will serve as an attractor. The report states that “We have not yet offered professor positions to spouses with significantly lower academic standing.” We suggest that IST develop a clear policy on this; obviously, there are tradeoffs, but we encourage a flexible approach.

The panel notes that indeed the market is very competitive with many top schools competing to get the best talent from the narrow pipeline of women available for academic careers. We encourage IST to continue to keep up their efforts which we hope will yield results going forward.

Noting that IST is likely to be more attractive to candidates in Europe (though we appreciate the more international nature of the faculty), we encourage IST to become a host or lead in creating a European version of the Rising Stars Workshop, modeled after an ongoing series in the US (see https://risingstars21-eecs.mit.edu for the most recent). The workshop would be focused on European universities as potential employers, but could attract potential candidates from anywhere in the world. This requires substantial effort and resources. However, we encourage IST to collaborate and coordinate with other universities such as ETH, EPFL, Oxford, UCL, Leuven, KTH, TU Munich, Sorbonne Université, etc., to establish such a program. Many of the faculty we met noted that IST’s strengths include a very good quality of life, a collegial atmosphere, and intellectually rewarding collaborations. By leading an event like Rising Stars at IST, the panel believes that IST can expose a bigger population of budding female scientists to

these strengths, leading them to consider IST as a compelling long-term career destination. Over time this kind of program could result in a robust pipeline for recruiting women.

A program that IST runs, STEM fatale, was highlighted in the documents shared with the panel. The panel felt the name did not serve the aim of increasing female representation in STEM through positive imagery. The phrase femme fatale (upon which the program name STEM fatale makes a pun) is often used in a derisory manner in modern sensibilities. [Googling the phrase femme fatale surfaces: “A femme fatale, sometimes called a maneater or vamp, is a stock character of a mysterious, beautiful, and seductive woman whose charms ensnare her lovers, often leading them into compromising, deadly traps.”] While we appreciate that the original naming was probably intended to be humorous, it will be off-putting to people from many cultures and further hurt IST’s ability to hire women from international backgrounds.

It is important to set up a strong mentoring network for women (for existing PhDs and postdocs and future faculty hires) so that women believe that the ecosystem at IST is supportive for their long-term careers.

2. Computer Science: Currently IST Austria’s faculty is mainly visible in theoretical computer science. What is the ideal balance between theorists and experimentalists / applied computer science? Is there a need to hire faculty in more applied fields?

Please refer to Q2 in the “General” section above.

3. How can IST Austria compete with industry to attract and retain faculty? What can we specifically offer that industry cannot offer?

Obviously, IST cannot compete on salary, given that total compensation for top researchers easily reaches seven figures at the leading industry labs. There are, nonetheless, many researchers who join top industry labs not for the money but for the opportunity to work on high-impact, cutting-edge research with plenty of resources and with very little overhead from teaching, administration, and proposal-writing.

IST, perhaps more than most universities, can offer some of the same advantages, as well as tenure and the promise that faculty will never be forced to work on projects driven solely by the corporation’s profit motive. For many industry labs, the “blue-sky basic research” period soon evaporates - even at Bell Labs - and none can match the longevity and independence of the research universities.

One aspect that perhaps could be enhanced, and that industry cannot offer, is the opportunity to do high-impact research *in the public interest*. IST needs to cultivate more long-term collaborations with external organizations. To pick just one local example, the CTBTO and IAEA in Vienna have many important computational, statistical, signal-processing, and computer

security problems related to their respective responsibilities. Having nearby R&D labs for major tech companies would also help with this aspect; at some universities, however, these labs are directed by current faculty on partial leave, which IST does not allow. Finally, if IST decides to grow in the direction of computational precision health, collaborations with nearby hospitals and medical schools will be essential.

Although the panel did not get into the details of IST's IP policy, it is our understanding that it works well for IP issues arising from faculty consulting but may be more problematic for faculty-led start-up companies. IST should view such companies not as deleterious to faculty research but as a legitimate vehicle for research impact and amplification of experimental capabilities.

4. Data science: Currently data science at IST Austria has a strong topical orientation (e.g. neuroscience, bioinformatics). Should this be balanced with more theoretical faculty (closer to statistics, artificial intelligence, etc.)?

We reviewed CS and data science faculty. Although we only saw a limited selection of both, it is apparent that the CS faculty as a whole form a quite coherent body, working in a highly effective manner in a somewhat restricted, theoretical portion of the overall domain of CS. This allows them to punch at and above their weight, helps with faculty and student recruitment, and allows them the benefit of disciplinary coherence. Although our review was restricted to ISS, something similar could perhaps be said for the renowned science foci of IST.

By contrast, the data science faculty, whilst individually excellent and world class, appear to be institutionally more precarious - and, in some cases, even uncertain about the reasons for their allocation to this 'area'. As IST expands, and the areas inevitably grow in importance as regularized groupings, this uncertainty might be exacerbated. At least one of the data scientists explicitly mentioned the need for colleagues with core expertise in statistical modelling and inference.

Even the wording of the question - the implicit implication that topical orientation is in opposition to theoretical rigour - reveals the uncertainty about the nature and worth of the *science* of data science. It is by now a given that the majority of the most exciting scientific discoveries and breakthroughs come at interdisciplinary boundaries, but the individuals on each side of the boundary are not service personnel for the others. This has yet to become properly commonplace when it is data science that is on one side of that boundary - particularly in institutions like IST, where the data scientists are sparse and thematically spread. Furthermore, in places where data science is vibrant, natural science and engineering students and postdocs gain substantial data analytical knowledge and skills, substantially boosting the benefits of this interdisciplinary interaction. Having one machine learning staff scientist to 'do' the machine learning for the natural science groups does not seem like a full solution to these issues.

Thus, although faculty working in statistics, machine learning and AI should definitely be a key part of any modern grouping in information sciences broadly construed, this would not be a substitute for a coherent plan for data science.