Patterns

Article

# Systematic analysis of 50 years of Stanford University technology transfer and commercialization

## Highlights

- Computational analysis of 4,512 inventions marketed by Stanford OTL since 1970
- Most profitable inventions are predominantly licensed by inventors' own startups
- Inventions involved larger teams over time
- Marketing abstracts predict future revenue of inventions

## Authors

Weixin Liang, Scott Elrod, Daniel A. McFarland, James Zou

## Correspondence

jamesz@stanford.edu

## In brief

Computational analysis of 4,512 inventions marketed by Stanford's Office of Technology Licensing between 1970 and 2020 characterizes how the academic innovation landscape has changed over time. We identified factors, such as the composition of the inventors, associated with the commercial success of the inventions. We also identified linguistic differences in how high- and low-revenue inventions in the same field are described and marketed.

Liang et al., 2022, Patterns *3*, 100584
September 9, 2022 
https://doi.org/10.1016/j.patter.2022.100584

CellPress

Patterns

CellPress
OPEN ACCESS

Article

# Systematic analysis of 50 years of Stanford University technology transfer and commercialization

Weixin Liang,[1] Scott Elrod,[2] Daniel A. McFarland,[3] and James Zou[1,4,5,*]
[1]Department of Computer Science, Stanford University, 350 Jane Stanford Way, Stanford, CA 94305, USA
[2]Office of Technology Licensing, Stanford University, Stanford, CA 94305, USA
[3]Graduate School of Education, Stanford University, Stanford, CA 94305, USA
[4]Department of Biomedical Data Science, Stanford University, 350 Jane Stanford Way, Stanford, CA 94305, USA
[5]Lead contact
*Correspondence: jamesz@stanford.edu
https://doi.org/10.1016/j.patter.2022.100584

**THE BIGGER PICTURE** Universities play an increasingly central role in research innovations and commercialization that drive technological development and economic growth. However, in-depth data science analysis of university technology transfer is underexplored in literature because the relevant data is often unavailable. To address this gap, we collaborated with the Stanford University Office of Technology Licensing (OTL) to curate a comprehensive dataset of all 4,512 inventions marketed by the OTL between 1970 to 2020. We have detailed information about each invention together with its generated revenue and cost, which critically captures outcomes missing in previous works. Examples of technologies licensed from Stanford include PageRank, recombinant DNA, and music synthesizers. Our study opens up a new perspective for analyzing the translation of research into practice and commercialization using large-scale computational and linguistics analysis.

1 2 3 4 5 **Concept:** Basic principles of a new data science output observed and reported

## SUMMARY

This article systematically investigates the technology licensing by Stanford University. We analyzed all the inventions marketed by Stanford's Office of Technology Licensing (OTL) between 1970 to 2020, with 4,512 inventions from 6,557 inventors. We quantified how the innovation landscape at Stanford changed over time and examined factors that correlate with commercial success. We found that the most profitable inventions are predominantly licensed by inventors' own startups, inventions have involved larger teams over time, and the proportion of female inventors has tripled over the past 25 years. We also identified linguistic features in how the inventors and OTL describe the inventions that significantly correlate with the invention's future revenue. Interestingly, inventions with more adjectives in their abstracts have worse net income. Our study opens up a new perspective for analyzing the translation of research into practice and commercialization using large-scale computational and linguistics analysis.

## INTRODUCTION

The role of American research universities has evolved and expanded in recent decades.[1] While the traditional mission of universities has long been to educate young people and to discover and transmit new disciplinary knowledge,[2] today, many universities have added technological invention and commercialization as part of their core mission.[3] This is manifested by changes in some universities' mission statements, the proliferation and enlargement of offices of technology licensing (OTLs),[4] the increase in the number of invention disclosures, patents, and licenses, and changes in tenure and promotion criteria to encourage the commercialization of university-generated knowledge.[5] As universities are playing an increasingly central role in technological inventions that drive economic growth,[6] many technology-reliant companies have reduced their budgets for internal research and development, opting to lean more heavily on collaborations with universities.[3,7–10]

With the increased emphasis on university technology transfer, many universities in the US have established OTLs.[4] An OTL serves as a mediator between the suppliers of innovations (university scientists) and those who can potentially commercialize them, i.e., industry,[11,12] centralizing university inventions and facilitating their commercialization through licensing to existing firms or startup companies of inventors. As examples, technologies that Stanford University's OTL has commercialized include the recombinant DNA technology that helped to jump start the biotechnology industry, internet search engines (e.g., Google PageRank), functional antibodies, and music synthesizers. The activities of OTLs have important economic and policy implications since licensing agreements and university-based startups can result in additional revenue for the university, employment opportunities, and local economic and technological spillovers through the stimulation of additional research and development (R&D) investment and job creation.[13–15] To incentivize university scientists, universities typically share licensing income with the inventors and the inventor's department.[16] For example, Stanford's royalty-sharing policy is to divide a third of the net income to the inventor, a third to the inventors' departments, and a third to the inventors' schools.

We focus our analysis on Stanford OTL because it is one of the most active and impactful technology transfer centers. Stanford OTL has long been regarded as a canonical approach for many universities in both the US and abroad.[17,18] Established in 1970, Stanford's OTL is one of the older OTLs.[4,19,20] Stanford has one of the most successful technology transfer programs, which has contributed to substantial commercial activity. According to the 2020 annual survey of the Association of University Technology Managers (AUTM),[4] Stanford ranks among the top five universities in the US across each of the key technology transfer performance metrics, including license income received, invention disclosures, US patents issued, and startup companies formed.

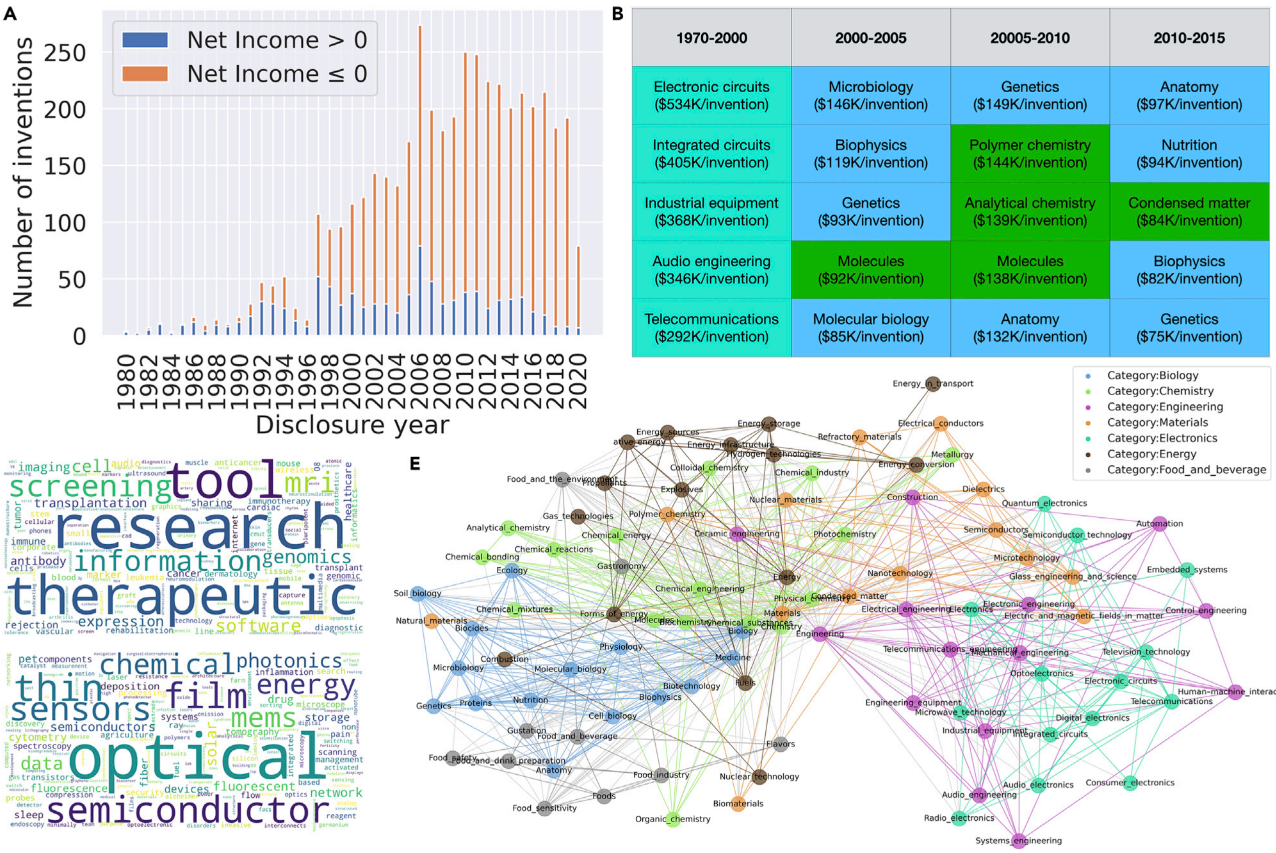
Despite the increasing importance of technology licensing, systematic data science analysis of university technology transfer is not common. Much of the previous research focuses on a few particular inventions at specific universities.[17,19,21,22] For example, Colyvas et al. investigated the early formation and the institutionalization process of Stanford's OTL.[17,19,22] Another line of research focuses on the determinants of licensing. Shane et al. analyzed early MIT inventions issued during the 1980–1996 period and found that university inventions are more likely to be licensed when patents are effective.[23] Using the same data, Dechenaux et al. explored the effect of appropriability on commercialization of inventions,[24] and Shane et al. found that new ventures with founders having direct and indirect relationships with venture investors are most likely to receive venture funding.[25] Huang et al. performed a systematic analysis of life sciences patents in MIT from 1983 to 2017. They include a number of outcome measurements that are unique to the biopharmaceutical industry, such as Orange Book citations, drug candidates discovered, and US drug approvals.[21] Other works focused on faculties' decisions on invention disclosure,[9,26–31] showing that faculty decisions to disclose are shaped by their perceptions of the benefits of patent protection[9] and the historical structure and mission of the university.[27]

Several works performed cross-university studies at a coarser granularity to measure the efficiency of university technology transfer.[32,33] For example, Thursby et al. found that the rise in university technology transfer is the result of a greater willingness of university researchers to patent their inventions and an increase in outsourcing of R&D by firms via licensing.[31,33] Subsequent work showed that higher percentages of royalty shares for faculty members,[26,34] and age of OTLs,[35,36] are associated with greater licensing income. Beyond the US, studies on the efficiency of university technology transfer have also been conducted in other countries including the UK,[37] Spain,[38] and Italy.[39,40] In sum, this line of research suggested that the key impediments to better university technology transfer performance tend to be organizational,[41] which includes incentives, relating both to pecuniary and non-pecuniary rewards, such as credit toward tenure and promotion, the staffing and compensation practices of the OTLs, university culture, milieu of entrepreneurism, and group norms.[13,42,43]

More broadly, previous research has investigated factors that drive scientific innovation. Although scientific innovation is widely accepted to be highly uncertain and unpredictable, previous research found that scientific projects that posit unexpected relationships between domains receive greater attention and are more richly rewarded than projects that explore more commonplace connections.[44–47] Although external factors such as the overall funding landscape and economic conditions could also affect scientific innovation, research teams are the engines of modern science.[48] The growth of prevalence and size of teams has been one of the most universal trends across all areas of scientific and scholarly investigation.[49] Prior experimental and observational studies reveal that demographic diversity benefits innovation.[50–52] Smaller teams have tended to disrupt science and technology with new ideas and opportunities, whereas larger teams have tended to develop existing ones.[53,54] Another line of research investigated patent-to-paper citations to assess the route from public research to economic and social impact, which highlights the importance of basic research and public research.[55–59]

Large-scale computational analysis of university technology transfer by OTLs has been limited in the literature. This gap motivates our comprehensive computational analysis leveraging the unprecedented data of 4,512 marketed inventions from 6,557 inventors at Stanford since the founding of its OTL in 1970. These are the inventions that Stanford's OTL prioritized for marketing over the 13,485 disclosed inventions during that period of time.

In our analysis, we focus on (1) quantifying how the innovation landscape at Stanford evolved over 50 years and (2) examining the factors that correlate with commercial success. We organize our analysis by first quantifying the holistic trends of invention at Stanford over time. We then analyze the inventors driving the innovations—their demographics, team composition, and the effects of licensing by inventor startups. Finally, a particularly interesting aspect of inventions is how they are publicly marketed, which is also a key role of the OTL. Therefore, we further analyze linguistic features in how an invention is presented in its title and abstract, as these semantic footprints enable us to gain insights into what the OTL believes are important to highlight. Our study opens up a new perspective for analyzing the




| 1970-2000 | 2000-2005 | 20005-2010 | 2010-2015 |
|---|---|---|---|
| Electronic circuits ($534K/invention) | Microbiology ($146K/invention) | Genetics ($149K/invention) | Anatomy ($97K/invention) |
| Integrated circuits ($405K/invention) | Biophysics ($119K/invention) | Polymer chemistry ($144K/invention) | Nutrition ($94K/invention) |
| Industrial equipment ($368K/invention) | Genetics ($93K/invention) | Analytical chemistry ($139K/invention) | Condensed matter ($84K/invention) |
| Audio engineering ($346K/invention) | Molecules ($92K/invention) | Molecules ($138K/invention) | Biophysics ($82K/invention) |
| Telecommunications ($292K/invention) | Molecular biology ($85K/invention) | Anatomy ($132K/invention) | Genetics ($75K/invention) |

**Figure 1. Overview of the Stanford inventions data**

(A) Number of inventions by year that Stanford's Office of Technology Licensing marketed. The color of the stacked bar chart indicates whether the cumulative net income (until June 31, 2021) is positive.

(B) Categories with the highest average net income across years. The numbers in parentheses indicate the average net income. Cell colors indicate the root category (teal: electronics, blue: biology, green: chemistry).

(C and D) Overrepresented keywords of (C) above-median income inventions (net income above the median for the same year) and (D) below-median income inventions. We identified words with the greatest log likelihood ratio of appearing in above-median invention keywords versus below-income invention keywords. The size of each term in the word cloud corresponds to its log likelihood ratio.

(E) Visualization of the sub-categories and the collaborative relationship among them. Each node represents a sub-category, and the edge is defined as the percentage of overlapping inventions that the two sub-categories share. Node color indicates the root category. Intra-category edges are colored using the color of the root category. Inter-category edges are colored gray.

translation of research into practice and commercialization by using fine-grained, large-scale computational and linguistic analysis.

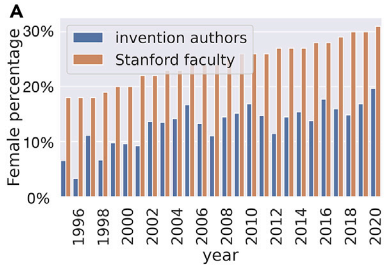
## RESULTS

### Overview of the Stanford inventions data

Our study leverages unprecedented data access to 4,512 marketed inventions, corresponding to 6,557 inventors from Stanford between 1970 and 2020, provided to us by the Stanford OTL for analysis. The number of marketed inventions increased rapidly from 1980 (4 inventions per year) to 2010 (250 inventions per year) and plateaued in the 2010s (Figure 1A). The rise of the internet greatly facilitated marketing, contributing to the large increase. Following the convention of the OTL, we use net income, which is defined as the total licensing income minus the cumulative expense (e.g., patent application and litigation costs) as a measure of the outcome of an invention. The total net income of the inventions for all years considered is $581 M, and the average net income is $0.13 M. Overall, most inventions have a negative net income, and only 20% of inventions in this dataset have produced positive net income (Figure 1A).

Each invention is assigned to one or more categories (e.g., "biophysics") and keywords (e.g., "Alzheimer disease") by the OTL. The categories with the highest average net income changed across the years (Figure 1B). Before 2000, the top net income categories were all in electronics, and after 2000, the top net income categories were in biology and chemistry. Since the net income is cumulative across time, recent inventions have a lower net income compared with older inventions because they had less time to accumulate income. We also identified the keywords that had the greatest log likelihood ratio of appearing in above-median inventions (net income above the median for the same year) versus below-median inventions (Figure 1C), and vice versa (Figure 1D). Words enriched in above-median income inventions tend to be life sciences terms, such as therapeutic

A




B

| Category | 1980-2000 | 2000-2005 | 20005-2010 | 2010-2015 | 2015-2020 |
|---|---|---|---|---|---|
| Biology | 2.47 | 2.88 | 3.08 | 3.24 | 3.29 |
| Chemistry | 2.51 | 2.73 | 3.05 | 3.22 | 3.25 |
| Engineering | 2.48 | 2.75 | 3.03 | 3.22 | 3.30 |
| Materials | 2.50 | 2.83 | 3.07 | 3.24 | 3.31 |
| Electronics | 2.58 | 2.65 | 2.91 | 3.22 | 3.48 |
| Energy | 2.88 | 3.11 | 3.21 | 3.23 | 3.23 |

Number of authors per invention

**Figure 2. Inventor demographics analysis**

(A) The percentage of female Stanford faculty and invention authors over the past 25 years.

(B) The number of authors per invention across different categories over time.

and genomics. In contrast, words enriched in low-income inventions tend to be associated with physical sciences, such as optical and photonics.

One invention can be assigned to multiple categories if it is relevant to different domains. For example, 17 medical imaging inventions disclosed in 2020 are assigned to both the radiology subcategory (under the biology category) and the computer vision subcategory (under the engineering category). Categories that co-occur in many inventions suggest that there is fruitful interdisciplinary research between them. We visualize the interaction relationship among different categories as a network (Figure 1E). There are substantial interactions between subcategories of biology and chemistry and between subcategories of engineering and electronics. Subcategories of materials science have diverse interactions with biology, chemistry, and engineering.

In all of the following analyses, to control for net income change over time, we use the net income rank, i.e., the normalized rank of the net income among the inventions with the same disclosure year. We also control the categorical difference by adding the categories as control variables in all linear regression analyses.

## Inventor demographics analysis

The proportion of female inventors has tripled from 6.5% in 1995 to 19.7% in 2020 (Figure 2A). The increase remains significant after controlling for categories ($p = 7.9E-07$; Table S5). However, despite such a rapid increase, overall, females are still underrepresented: the percentage of female inventors is consistently lower than the percentage of female faculty at Stanford by a large margin. For example, in 2019, the percentage of female faculty at Stanford was 30%, while only 20% of the inventors were female. A caveat here is that certain disciplines (e.g., medicine, engineering) can be more likely to file inventions than other disciplines.[16]

Beyond involving more females, inventions also involved larger teams over time. For example, the average number of inventors per invention under the biology category increased from 2.47 in 1980–2000 to 3.29 in 2015–2020 (Figure 2B). Such an increase is consistent across different categories ($p = 1.7E-30$; Table S6), indicating that the invention environment at Stanford is increasingly collaborative. In addition, we found that the inventions from teams of only first-time inventors have a higher net income than other inventions ($p = 3.1E-02$; Table S7). This highlights the importance of being open to first-time inventors.

## Self-licensing

Stanford's entrepreneurial culture is also reflected in our data. Around 20% of the inventions were licensed by the inventor's own startups, which we refer to as "self-licensing." Overall, the self-licensing rate increases over time ($p = 4.0E-02$; Figure 3A; Table S8). The interesting peak of the self-licensing rate in 1995–1999 might be related to the dot-com bubble. We also found that inventions with high net income are predominantly self-licensing inventions (Figure 3B). For example, all inventions that have generated more than \$10 M net income are self-licensed, and the self-licensing rate for the inventions with \$1–\$10 M net income is 59%. In contrast, the self-licensing rate for inventions with less than \$10 K net income is 16%. After controlling for categories and years, self-licensing is still strongly associated with higher net income ($p = 4.1E-08$; Table S9). This finding is consistent with previous research showing that startups with direct connections to the university tend to be more successful than otherwise similar startups.[60] In addition, the self-licensing rate is higher in the biology category ($p = 3.0E-05$) and lower in the electronics category ($p = 6.4E-04$).

## Linguistic analysis on OTL marketing

An important role of the Stanford OTL is to market the researchers' inventions to potentially interested companies. Marketing is typically initiated through a marketing abstract created by the OTL that describes the invention to the public. Therefore, to gain insights into OTL marketing, we analyze two main questions. (1) How have marketing abstracts changed over the years? (2) Which linguistic features in the marketing abstracts are associated with the commercial outcome of the invention? Similar text analysis techniques have been applied to scientific innovation studies in the literature.[50]

The marketing abstracts have changed substantially over the years. The average length of the marketing abstracts has nearly doubled: from 144 words in 1980–1990 to 241 words in 2015–2020 (Figure 4A). The increase remains statistically significant after controlling for categories ($p = 7.7E-42$; Tables S1 and S2). Interestingly, the titles of the marketing abstracts are also getting longer ($p = 6.3E-19$; Table S1) and have 10× more adjectives ($p = 2.4E-52$; Figure 4B; Table S3) from 1980–1990 (1%) to 2015–2020 (12%). This might suggest that inventions are becoming increasingly specialized, which would require longer text and more adjectives to describe them.

Beyond the temporal changes, we also identified linguistic features in how the OTL describes the inventions that significantly correlate with the invention's future revenue. We found that inventions with longer marketing abstracts ($p = 2.2E-04$) or more adjectives in the marketing abstracts ($p = 1.4E-05$) are associated with worse net income (Table S4). Interestingly, we found that words like "novel" ($p = 3.57e-08$), "significant" ($p = 2.00e-03$), and "effective" ($p = 8.51e-03$) correlate negatively with the net income, even after controlling for categories. These adjectives remain statistically significant after adjusting for multiple

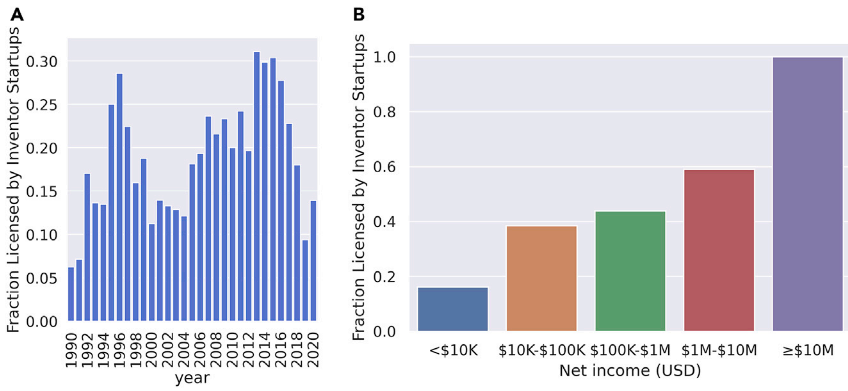



**Figure 3. Self-licensing (inventions licensed by the inventor's own startups)**

(A) The fraction of inventions licensed by inventor startups over time.

(B) The fraction of inventions in each net income group that the inventors license. The sample sizes for each net income category are: <\$10 K: 3,776 inventions; \$10–\$100 K: 465 inventions; \$100 K–\$1 M: 212 inventions; \$1–\$10 M: 56 inventions; ≥\$10 M: 5 inventions.

hypothesis testing with a false discovery rate of 0.05. In contrast, only a few adjectives correlate positively with net income.

Analyses of inventors' abstracts show similar results: The length of inventors' abstracts has also substantially increased over time ($p = 6.3E-19$) after controlling for categories (Table S1). In addition, both the length of the inventors' abstracts ($p = 1.3E-02$) and the fraction of adjectives ($p = 2.7E-05$) correlate negatively with net income (Table S4). We further investigated the correlation between net income and the usage of each adjective in the inventors' abstracts (Figure S1A). Similarly, we found that in inventor's abstracts, adjectives like ''significant'' ($p = 4.04e-04$), ''novel'' ($p = 2.38e-02$), and ''effective'' ($p = 3.27e-03$) also correlate negatively with the net income, even after controlling for categories. These adjectives also remain statistically significant after adjusting for multiple hypothesis testing with a false discovery rate of 0.05. One possible explanation is that for more incremental inventions, inventors tend to write longer abstracts and use more adjectives to highlight their novelties and advantages over existing technologies, and the writing of marketing abstracts by OTLs might be influenced by the inventors' abstracts.

Finally, to quantify the distinction between the marketing abstract of above-median income inventions and those of low-income inventions, we trained machine-learning classifiers to predict whether the net income rank is above 0.5, i.e., whether the net income is above the median for the same disclosure year (Figure 4D). A classifier using term frequency-inverse document frequency (TF-IDF) features achieves a 0.71 area under the receiver operating characteristic (AUROC) on the hold-out test set. The BERT classifier, which utilizes deep learning to provide contextual features for each word, is highly accurate, with a 0.76 AUROC score. In contrast, the baseline classifier that takes category annotations as input only achieves a low 0.57 AUROC score, suggesting that the linguistic patterns we identify here are not driven by different styles of presenting different categories of inventions. Experiments on inventors' abstracts show similar results (Figure S1B). This suggests that the abstracts of above-median income inventions have clearly distinguishing textual features beyond categorical differences.

## DISCUSSION

This paper provides a systematic and quantitative characterization of the technology licensing pipeline at Stanford between 1970 and 2020, with 4,512 inventions from 6,557 inventors. Our analysis characterizes how the innovation landscape at Stanford changed over time: the top-income invention categories shifted from electronics to life sciences after 2000. The inventions might also be increasingly specialized, as indicated by the substantial increase in the length of both the titles and the abstracts for describing them.

Our demographic analysis suggests that inventions involved larger teams over time across all categories. The proportion of female inventors has tripled over the past 25 years, though they are still underrepresented. This finding is consistent with previous research findings on the gender gap in patenting.[61] Proactive efforts can be taken to support diverse faculties in translating their research to industry. Our analysis also highlights the important role of inventors in commercializing research: the most profitable inventions are predominantly licensed by inventors' own startups, and such self-licensing practices are also becoming increasingly popular over time. This finding is consistent with previous research showing that startups with direct connections to the university tend to be more successful than otherwise similar startups.[60] Several other papers have also shown evidence for a positive relationship between faculty involvement and commercialization outcomes.[23,25,62] Overall, the self-licensing rate increases over time, and there is an interesting peak of the self-licensing rate in 1995–1999 that might be related to the dot-com bubble.

An important role of the Stanford OTL is to market the researchers' inventions to potentially interested companies. A primary way of this marketing is through the OTL providing a marketing abstract that describes the invention to the public. Our linguistic analysis identified linguistic features in how the OTL describes the inventions that significantly correlate with the invention's future revenue. Interestingly, inventions with more adjectives in the marketing abstracts are associated with worse net income. Adjectives like ''novel,'' ''effective,'' and ''significant'' in the marketing abstracts correlate negatively with the net income, even after controlling for categories and year. One possible explanation is that for more incremental inventions, inventors tend to write longer abstracts and use more adjectives to highlight their novelties and advantages over existing technologies, and the writing of marketing abstracts by OTL might be influenced by the inventors' abstracts. Furthermore, the strong predictive performance at discriminating both the author and marketing abstracts of inventions with above-median versus below-median income, after controlling for categories, exemplifies their substantial linguistic

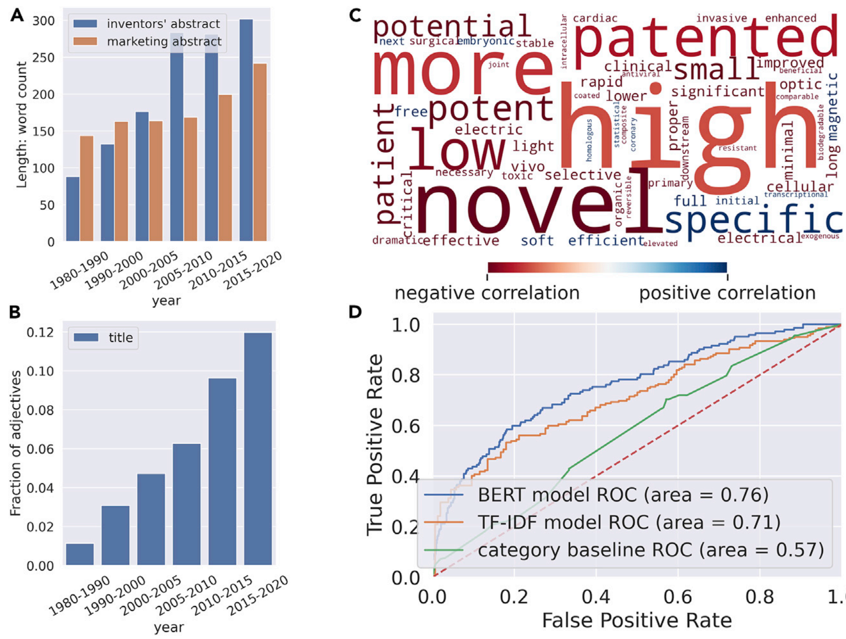



**Figure 4. Linguistic analysis on OTL marketing**

(A) The average length of OTL marketing abstracts and inventors' abstracts over time.

(B) The average fraction of adjectives in titles over time.

(C) The correlation between the occurrence of each adjective in the marketing abstract and net income rank. Shown here are adjectives with p <0.05. Font size indicates the frequency of the word. Text color indicates the correlation coefficient with net income rank after controlling for categories: red indicates negative correlation, and blue indicates positive correlation.

(D) Machine-learning classifiers with the marketing abstracts as inputs to predict whether the net income of an invention will be above the median net income of the inventions of the same disclosure year. TF-IDF, the classifier using term frequency-inverse document frequency features; BERT, the state-of-the-art text classifier that utilizes deep learning to provide contextual features for each word. Category baseline: only using category tags of each invention as inputs. Shown are receiver operating characteristic (ROC) curves on the hold-out test set. A classifier using TF-IDF features achieves a 0.71 area under the receiver operating characteristic (AUROC) on the hold-out test set.

difference and opens up new possibilities for further linguistic analysis. Future works include incorporating further semantic analysis on the abstracts to measure the scientific novelty of the inventions.[50]

The findings of this study have to be considered in light of some limitations. First, the invention licensing via OTLs represents only one facet of the transfer of technology from university to industry, though it is an important facet.[63] Second, we primarily focus on net income as the outcome metric because it is straightforward to quantify and is a key metric of OTL's own assessment. However, it is important to note that licensing income does not completely capture impact, and pursuing licensing income is not the ultimate goal of the Stanford OTL. The third limitation concerns the observational nature of our study. Although we have been careful in controlling for confounders like category and year in our statistical models, the results should not be interpreted as causal but rather as statistical associations. Finally, while our data focus on a single university, Stanford University, this is an important case study because Stanford is a leading center of innovation. Our findings also provide insights into the academic-industry partnership of Silicon Valley since many technologies and startups from Stanford are commercialized there. More work is needed to study the technology licensing at other universities with different entrepreneurial environments.

## EXPERIMENTAL PROCEDURES

### Resource availability

#### *Lead contact*

Further information and requests for code and data should be directed to and will be fulfilled by the lead contact, James Zou (jamesz@stanford.edu).

#### *Materials availability*

This study did not generate any physical materials.

#### *Data and code availability*

Aggregated data reported in this paper will be shared by the lead contact upon request. All original code has been deposited at Zenodo under https://doi.org/10.5281/zenodo.6959366 and is publicly available as of the date of publication. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

### Materials and methods

#### *Stanford inventions data*

Metadata for the subset of 4,512 inventions that were prioritized for marketing, corresponding to 6,557 inventors from Stanford between 1980 and 2020, were provided to us by the Stanford OTL for analysis. Many of the inventions were web marketed, which partly explains the rapid increase in the number of inventions in the 1990s. The OTL receives invention disclosures from Stanford faculty, staff, and students. Generally, faculty notify the OTL of their invention discoveries and delegate to the university all rights to negotiate licenses on their behalf.[64] After receiving the invention disclosures, the OTL evaluates the commercial potential of the invention and, if it is a patentable subject matter, decides whether to file a patent.[42] Our dataset contains inventions protected by both patents and copyright. For each invention, we have data on the title, name of inventors, the abstract, keywords, category tags, and disclosure date. We also have access to the cumulative revenue and the cumulative expense, from which we can derive the cumulative net income of each invention, which provides a measure of the impact of each invention. The net income in our dataset is calculated before sharing it with inventors and inventors' departments and schools. Given that we aimed to focus on the impact of each invention, we considered each invention (i.e., docket) as our unit of analysis. For more information about the Stanford OTL, we refer interested readers to https://otl.stanford.edu/. For examples of Stanford inventions, we refer interested readers to http://techfinder.stanford.edu/

#### *Categorization for inventions*

The original Stanford invention dataset contains the category tags only for a subset of 1,700 inventions, annotated by a third-party marketing platform. Therefore, we trained a machine-learning model to propagate category annotations for all inventions. The input of the categorization model is the keywords and the title of each invention. For each category tag, we trained a binary classifier by fine-tuning a BERT deep neural language model [65] to predict whether an invention belongs to this category or not. Using a held-out test set, we found that our categorization models achieve high classification performance: for most of the category tags, the categorization model has an AUC larger than 0.8. The AUC score is larger than 0.7 for all category tags. The categorization model is implemented using PyTorch 1.4.0. Since the categorization model we trained was highly accurate (AUC > 0.8), we used the full samples (4,512 inventions) along with the predicted category tags for our analyses.

#### Statistical analysis and multiple hypothesis testing

Data processing, statistical testing, and visualization were performed using Python v.3.7. We conducted a series of statistical tests using a Python-package statsmodel (https://www.statsmodels.org/stable/index.html). We supply p values as a tool for interpretation; we maintain the convention of 0.05 as the threshold for statistical significance. We performed the Benjamini-Hochberg procedure for multiple hypothesis testing with a false discovery rate of alpha = 0.05 using the statsmodel package. Our plots were generated using the matplotlib Python package.

#### Linguistic analysis: Predicting net income from both the author and marketing abstracts

We hypothesized that there are linguistic differences in how the inventors and OTLs describe the inventions. We performed linguistic analysis on inventors' abstracts (the abstracts of the invention disclosures written by university scientists), marketing abstracts (the abstracts rewritten by the Stanford OTL's marketing team for the audience of business and legal professionals), and the final invention titles edited by the Stanford OTL's marketing team. We split the author and marketing abstracts under consideration (inventions between 1980 and 2020) in an 80%/20% ratio as the train/test splits. We used the sklearn python package and trained a TF-IDF featurizer on the training data and then featurized both training and test data. Finally, we trained a logistic regression model based on the features. The AUC-ROC curve was evaluated on the test set. Furthermore, we also experimented with the BERT deep neural language model, implemented using PyTorch 1.4.0.

#### Linguistic analysis: Adjectives

Part of speech tagging provides the functionality of marking a word in the text to a particular part of speech (e.g., adjectives, nouns, pronouns, verbs) based on both its context and definition. We used the Python Package Spacy (https://spacy.io/), an industrial-strength natural language processing toolkit, to perform part of speech tagging and identify adjectives.

#### Demographic analysis: Stanford faculty gender data

We used the official number of the percentage of female faculty at Stanford over the years from the Faculty Demographics reports authored by the Stanford Office of Faculty Development, Diversity, and Engagement, which is publicly available at https://facultydevelopment.stanford.edu/data-reports/faculty-demographics

We follow the method used by previous research[66,67] for gender identification from names. This method has recently been validated on a dataset of scientist names extracted from the WoS database[68]. The gender of each reviewer and reviewing editor is inferred from their names using a Python-package gender-guesser (https://pypi.python.org/pypi/gender-guesser/). Previous research shows that the gender-guesser package achieves the lowest misclassification rate and minimizes bias[68]. The validation performed by Santamaría and Mihaljević (2019)[68] limited misclassification to 1.5% for European names, 3.6% for African names, and 6.4% for Asian names.

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.patter.2022.100584.

## ACKNOWLEDGMENTS

We thank I. Kumar, A. Manjunath, J. Yang, Y. Li, K. Immergluck, and members of the Stanford OTL for feedback.

## AUTHOR CONTRIBUTIONS

Designed research, W.L., S.E., and J.Z.; analyzed data, W.L.; performed research, all authors; wrote the paper, W.L. and J.Z.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 3, 2022
Revised: July 17, 2022
Accepted: August 10, 2022
Published: September 1, 2022

## REFERENCES

1. Mowery, D.C., Nelson, R.R., Sampat, B.N., and Ziedonis, A.A. (2015). Ivory Tower and Industrial Innovation: University-Industry Technology Transfer before and after the Bayh-Dole Act (Stanford University Press).
2. Altbach, P.G., Gumport, P.J., and Berdahl, R.O. (2005). American Higher Education in the Twenty-First Century: Social, Political, and Economic Challenges (JHU Press).
3. Biancani, S., Dahlander, L., McFarland, D.A., and Smith, S. (2018). Superstars in the making? The broad effects of interdisciplinary centers. Res. Policy *47*, 543–557. https://doi.org/10.1016/j.respol.2018.01.014.
4. AUTM (2020). AUTM US Licensing Activity Survey: 2020 (Association of University Technology Managers). (AUTM)).
5. Mowery, D.C., Sampat, B.N., and Ziedonis, A.A. (2002). Learning to patent: institutional experience, learning, and the characteristics of U.S. university patents after the bayh-dole act, 1981-1992. Manage. Sci. *48*, 73–89. https://doi.org/10.1287/mnsc.48.1.73.14278.
6. Goldstein, H.A. (2010). The 'entrepreneurial turn' and regional economic development mission of universities. Ann. Reg. Sci. *44*, 83–109. https://doi.org/10.1007/s00168-008-0241-z.
7. Geiger, R.L., and Sá, C.M. (2008). Tapping the Riches of Science: Universities and the Promise of Economic Growth (Harvard University Press).
8. Salter, A.J., and Martin, B.R. (2001). The economic benefits of publicly funded basic research: a critical review. Res. Policy *30*, 509–532. https://doi.org/10.1016/S0048-7333(00)00091-3.
9. Owen-Smith, J., and Powell, W.W. (2001). To patent or not: faculty decisions and institutional success at technology transfer. J. Technol. Transf. *26*, 99–114. https://doi.org/10.1023/A:1007892413701.
10. Powell, W., White, D., Koput, K., and Owen-Smith, J. (2005). Network dynamics and field evolution: the growth of Interorganizational collaboration in the life sciences. Am. J. Sociol. *110*, 1132–1205. https://doi.org/10.1086/421508.
11. Bradley, S., Hayter, C.S., and Link, A. (2013). Models and Methods of University Technology Transfer (Social Science Research Network).
12. Yang, J., Chesbrough, H., and Hurmelinna-Laukkanen, P. (2022). How to appropriate value from general-purpose technology by applying open innovation. Calif. Manage. Rev. *64*, 24–48. https://doi.org/10.1177/00081256211041787.
13. Siegel, D.S., Veugelers, R., and Wright, M. (2007). Technology transfer offices and commercialization of university intellectual property: performance and policy implications. Oxf. Rev. Econ. Policy *23*, 640–660. https://doi.org/10.1093/oxrep/grm036.
14. Siegel, D.S., and Phan, P.H. (2005). Analyzing the effectiveness of university technology transfer: implications for entrepreneurship education. In University Entrepreneurship and Technology Transfer, G.D. Libecap, ed. (Emerald Group Publishing Limited), pp. 1–38.
15. Bozeman, B. (2000). Technology transfer and public policy: a review of research and theory. Res. Policy *29*, 627–655. https://doi.org/10.1016/S0048-7333(99)00093-1.
16. Thursby, J.G., Jensen, R., and Thursby, M.C. (2001). Objectives, characteristics and outcomes of university licensing: a survey of Major. J. Technol. Transf. *26*, 59–72. https://doi.org/10.1023/A:1007884111883.
17. Colyvas, J.A., and Powell, W.W. (2006). Roads to institutionalization: the remaking of boundaries between public and private science. Res.

Organ. Behav. 27, 305–353. https://doi.org/10.1016/S0191-3085(06)27008-4.

18. Leute, K. (2005). Patenting and licensing of university-based genetic inventions–a view from experience at stanford university's office of technology licensing. Community Genet. 8, 217–222. https://doi.org/10.1159/000087958.
19. Colyvas, J.A. (2007). From divergent meanings to common practices: the early institutionalization of technology transfer in the life sciences at Stanford University. Res. Policy 36, 456–476. https://doi.org/10.1016/j.respol.2007.02.019.
20. Mowery, D.C., Nelson, R.R., Sampat, B.N., and Ziedonis, A.A. (2001). The growth of patenting and licensing by U.S. universities: an assessment of the effects of the Bayh–Dole act of 1980. Res. Policy 30, 99–119. https://doi.org/10.1016/S0048-7333(99)00100-6.
21. Huang, S., Siah, K.W., Vasileva, D., Chen, S., Nelsen, L., and Lo, A.W. (2021). Life sciences intellectual property licensing at the Massachusetts Institute of Technology. Nat. Biotechnol. 39, 293–301. https://doi.org/10.1038/s41587-021-00843-5.
22. Colyvas, J.A., and Powell, W.W. (2007). From vulnerable to venerated: the institutionalization of academic entrepreneurship in the life sciences. In The Sociology of Entrepreneurship, M. Ruef and M. Lounsbury, eds. (Emerald Group Publishing Limited), pp. 219–259.
23. Shane, S. (2002). Selling university technology: patterns from MIT. Manage. Sci. 48, 122–137. https://doi.org/10.1287/mnsc.48.1.122.14281.
24. Dechenaux, E., Goldfarb, B., Shane, S., and Thursby, M. (2008). Appropriability and commercialization: evidence from MIT inventions. Manage. Sci. 54, 893–906. https://doi.org/10.1287/mnsc.1070.0780.
25. Shane, S., and Stuart, T. (2002). Organizational endowments and the performance of university start-ups. Manage. Sci. 48, 154–170. https://doi.org/10.1287/mnsc.48.1.154.14280.
26. Friedman, J., and Silberman, J. (2003). University technology transfer: do incentives, Management, and Location matter? J. Technol. Transf. 28, 17–30. https://doi.org/10.1023/A:1021674618658.
27. Feldman, M., and Desrochers, P. (2003). Research universities and local economic development: Lessons from the history of the Johns Hopkins university. Ind. Innov. 10, 5–24. https://doi.org/10.1080/1366271032000068078.
28. Jensen, R.A., Thursby, J.G., and Thursby, M.C. (2003). Disclosure and licensing of University inventions: 'The best we can do with the s**t we get to work with. Int. J. Ind. Organ. 21, 1271–1300. https://doi.org/10.1016/S0167-7187(03)00083-3.
29. Markman, G.D., Gianiodis, P.T., Phan, P.H., and Balkin, D.B. (2005). Innovation speed: Transferring university technology to market. Res. Policy 34, 1058–1075. https://doi.org/10.1016/j.respol.2005.05.007.
30. Roberts, E.B., and Peters, D.H. (1981). Commercial innovation from university faculty. Res. Policy 10, 108–126. https://doi.org/10.1016/0048-7333(81)90001-9.
31. Thursby, J.G., and Thursby, M.C. (2002). Who is selling the ivory tower? sources of growth in university licensing. Manage. Sci. 48, 90–104. https://doi.org/10.1287/mnsc.48.1.90.14271.
32. Anderson, T.R., Daim, T.U., and Lavoie, F.F. (2007). Measuring the efficiency of university technology transfer. Technovation 27, 306–318. https://doi.org/10.1016/j.technovation.2006.10.003.
33. Thursby, J.G., and Kemp, S. (2002). Growth and productive efficiency of university intellectual property licensing. Res. Policy 31, 109–124. https://doi.org/10.1016/S0048-7333(00)00160-8.
34. Lach, S., and Schankerman, M. (2004). Royalty sharing and technology licensing in universities. J. Eur. Econ. Assoc. 2, 252–264. https://doi.org/10.1162/154247604323067961.
35. Carlsson, B., and Fridh, A.-C. (2002). Technology transfer in United States universities. J. Evol. Econ. 12, 199–232. https://doi.org/10.1007/s00191-002-0105-0.
36. Powers, J.B., and McDougall, P.P. (2005). University start-up formation and technology licensing with firms that go public: a resource-based view of academic entrepreneurship. J. Bus. Ventur. 20, 291–311. https://doi.org/10.1016/j.jbusvent.2003.12.008.
37. Chapple, W., Lockett, A., Siegel, D., and Wright, M. (2005). Assessing the relative performance of U.K. university technology transfer offices: parametric and non-parametric evidence. Res. Policy 34, 369–384. https://doi.org/10.1016/j.respol.2005.01.007.
38. Lafuente, E., and Berbegal-Mirabent, J. (2019). Assessing the productivity of technology transfer offices: an analysis of the relevance of aspiration performance and portfolio complexity. J. Technol. Transf. 44, 778–801. https://doi.org/10.1007/s10961-017-9604-x.
39. Battaglia, D., Landoni, P., and Rizzitelli, F. (2017). Organizational structures for external growth of University Technology Transfer Offices: an explorative analysis. Technol. Forecast. Soc. Change 123, 45–56. https://doi.org/10.1016/j.techfore.2017.06.017.
40. Muscio, A. (2010). What drives the university use of technology transfer offices? Evidence from Italy. J. Technol. Transf. 35, 181–202. https://doi.org/10.1007/s10961-009-9121-7.
41. Macho-Stadler, I., Pérez-Castrillo, D., and Veugelers, R. (2007). Licensing of university inventions: the role of a technology transfer office. Int. J. Ind. Organ. 25, 483–510. https://doi.org/10.1016/j.ijindorg.2006.06.001.
42. Siegel, D.S., Waldman, D., and Link, A. (2003). Assessing the impact of organizational practices on the relative productivity of university technology transfer offices: an exploratory study. Res. Policy 32, 27–48. https://doi.org/10.1016/S0048-7333(01)00196-2.
43. Siegel, D.S., Waldman, D.A., Atwater, L.E., and Link, A.N. (2003). Commercial knowledge transfers from universities to firms: improving the effectiveness of university–industry collaboration. J. High Technol. Manag. Res. 14, 111–133. https://doi.org/10.1016/S1047-8310(03)00007-5.
44. Fleming, L. (2001). Recombinant uncertainty in technological search. Manage. Sci. 47, 117–132. https://doi.org/10.1287/mnsc.47.1.117.10671.
45. Foster, J.G., Rzhetsky, A., and Evans, J.A. (2015). Tradition and innovation in scientists' research Strategies. Am. Sociol. Rev. 80, 875–908. https://doi.org/10.1177/0003122415601618.
46. Uzzi, B., Mukherjee, S., Stringer, M., and Jones, B. (2013). Atypical combinations and scientific impact. Science 342, 468–472. https://doi.org/10.1126/science.1240474.
47. Fleming, L., and Sorenson, O. (2004). Science as a map in technological search. Strateg. Manag. J. 25, 909–928. https://doi.org/10.1002/smj.384.
48. Wuchty, S., Jones, B.F., and Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. science 316, 1036–1039. https://doi.org/10.1126/science.1136099.
49. Guimerà, R., Uzzi, B., Spiro, J., and Amaral, L.A.N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. Science 308, 697–702. https://doi.org/10.1126/science.1106340.
50. Hofstra, B., Kulkarni, V.V., Munoz-Najar Galvez, S., He, B., Jurafsky, D., and McFarland, D.A. (2020). The diversity-innovation Paradox in science. Proc. Natl. Acad. Sci. USA. 117, 9284–9291. https://doi.org/10.1073/pnas.1915378117.
51. Nielsen, M.W., Alegria, S., Börjeson, L., Etzkowitz, H., Falk-Krzesinski, H.J., Joshi, A., Leahey, E., Smith-Doerr, L., Woolley, A.W., and Schiebinger, L. (2017). Opinion: gender diversity leads to better science. Proc. Natl. Acad. Sci. USA. 114, 1740–1742. https://doi.org/10.1073/pnas.1700616114.
52. Bell, S.T., Villado, A.J., Lukasik, M.A., Belau, L., and Briggs, A.L. (2011). Getting specific about demographic diversity variable and team performance relationships: a Meta-analysis. J. Manag. 37, 709–743. https://doi.org/10.1177/0149206310365001.
53. Wu, L., Wang, D., and Evans, J.A. (2019). Large teams develop and small teams disrupt science and technology. Nature 566, 378–382. https://doi.org/10.1038/s41586-019-0941-9.
54. Christensen, C.M., Raynor, M.E., Dyer, J., and Gregersen, H. (2011). Disruptive Innovation: The Christensen Collection (The Innovator's Dilemma, the Innovator's Solution, the Innovator's DNA, and Harvard

Business Review Article 'How Will You Measure Your Life?') (4 Items) (Harvard Business Press).

55. Manjunath, A., Li, H., Song, S., Zhang, Z., Liu, S., Kahrobai, N., Gowda, A., Seffens, A., Zou, J., and Kumar, I. (2021). Comprehensive analysis of 2.4 million patent-to-research citations maps the biomedical innovation and translation landscape. Nat. Biotechnol. *39*, 678–683. https://doi.org/10.1038/s41587-021-00940-5.

56. Jefferson, O.A., Jaffe, A., Ashton, D., Warren, B., Koellhofer, D., Dulleck, U., Ballagh, A., Moe, J., DiCuccio, M., Ward, K., et al. (2018). Mapping the global influence of published research on industry and innovation. Nat. Biotechnol. *36*, 31–39. https://doi.org/10.1038/nbt.4049.

57. Ahmadpoor, M., and Jones, B.F. (2017). The dual frontier: patented inventions and prior scientific advance. Science *357*, 583–587. https://doi.org/10.1126/science.aam9527.

58. Huang, M.-H., Huang, W.-T., and Chen, D.-Z. (2014). Technological impact factor: an indicator to measure the impact of academic publications on practical innovation. J. Informetr. *8*, 241–251. https://doi.org/10.1016/j.joi.2013.12.004.

59. Ke, Q. (2020). An analysis of the evolution of science-technology linkage in biomedicine. J. Informetr. *14*, 101074. https://doi.org/10.1016/j.joi.2020.101074.

60. Siegel, D.S., and Wessner, C. (2012). Universities and the success of entrepreneurial ventures: evidence from the small business innovation research program. J. Technol. Transf. *37*, 404–415. https://doi.org/10.1007/s10961-010-9186-3.

61. Ding, W.W., Murray, F., and Stuart, T.E. (2006). Gender differences in patenting in the academic life sciences. Science *313*, 665–667. https://doi.org/10.1126/science.1124832.

62. Rothaermel, F.T., Agung, S.D., and Jiang, L. (2007). University entrepreneurship: a taxonomy of the literature. Ind. Corp. Change *16*, 691–791. https://doi.org/10.1093/icc/dtm023.

63. Hayter, C.S., Rasmussen, E., and Rooksby, J.H. (2020). Beyond formal university technology transfer: innovative pathways for knowledge exchange. J. Technol. Transf. *45*, 1–8. https://doi.org/10.1007/s10961-018-9677-1.

64. Litan, R.E., Mitchell, L., and Reedy, E.J. (2007). Commercializing university innovations: alternative approaches. Innovat. Pol. Econ. *8*, 31–57. https://doi.org/10.1086/ipe.8.25056198.

65. Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT *1*, 4171–4186.

66. Larivière, V., Ni, C., Gingras, Y., Cronin, B., and Sugimoto, C.R. (2013). Bibliometrics: Global gender disparities in science. Nature *504*, 211–213.

67. Helmer, M., Schottdorf, M., Neef, A., and Battaglia, D. (2017). Gender bias in scholarly peer review. Elife *6*, e21718.

68. Santamaría, L., and Mihaljević, H. (2018). Comparison and benchmark of name-to-gender inference services. PeerJ Comput. Sci. *4*, e156.

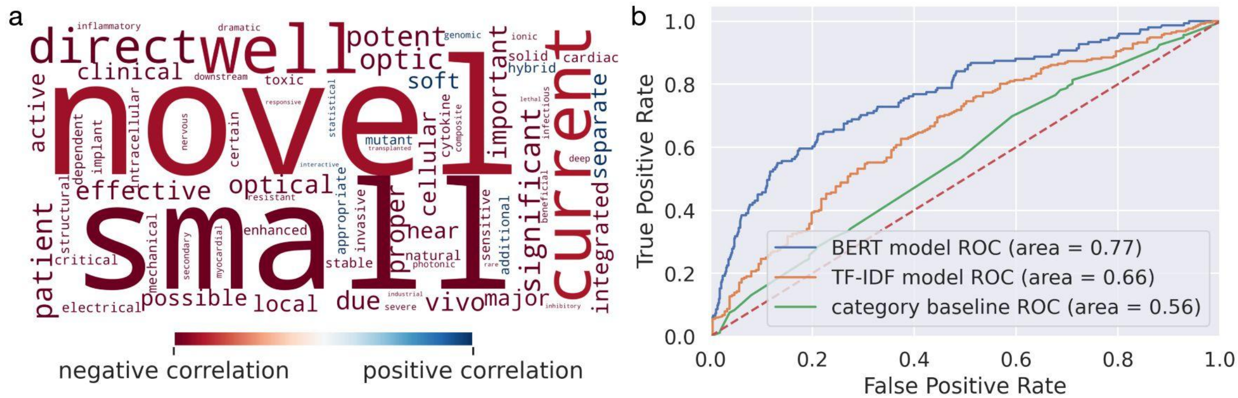
**Patterns, Volume *3***

## Supplemental information

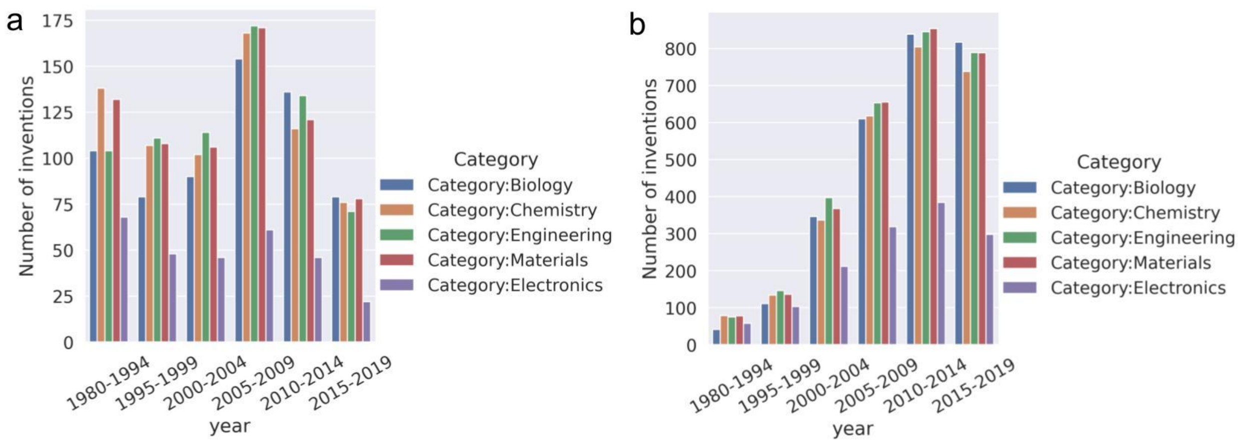
# Systematic analysis of 50 years of Stanford University technology transfer and commercialization

**Weixin Liang, Scott Elrod, Daniel A. McFarland, and James Zou**

# Supplementary Figures




**Supplementary Figure 1: Linguistic Analysis on OTL marketing (inventors' abstract).** (a) The correlation between the occurrence of each adjective in inventors' abstract and net income rank. Shown here are adjectives with P-value<0.05. Font size indicates the frequency of the word. Text color indicates the correlation coefficient with net income rank after controlling for categories: red indicates negative correlation and blue indicates positive correlation. (b) Machine learning classifiers with inventors' abstracts as inputs to predict whether the net income of an invention will be above the median net income of the inventions of the same disclosure year. TF-IDF: the classifier using term frequency-inverse document frequency features. BERT: the state-of-the-art text classifier that utilizes deep learning to provide contextual features for each word. Category baseline: only using category tags of each invention as inputs. Shown are receiver operating characteristic (ROC) curves on the hold-out test set. A classifier using term frequency-inverse document frequency (TF-IDF) features achieves an 0.66 area under the receiver operating characteristic (AUROC) on the hold-out test set.




**Supplementary Figure 2: Number of inventions by year by categories marketed by Stanford's Office of Technology Licensing.** (a) The Inventions that have a positive cumulative net income (until Jun 31, 2021). (b) The Inventions that have a non-positive cumulative net income (until Jun 31, 2021).

# Supplementary Tables

| DV: text length | | year | Category: Biology | Category: Chemistry | Category: Engineering | Category: Materials | Category: Electronics | Category: Energy |
|---|---|---|---|---|---|---|---|---|
| **Title** | Coefficient | 0.0773 | -0.0355 | 0.5398 | 0.1001 | 0.2816 | -0.6283 | -0.1138 |
| | P-value | 6.3E-19*** | 8.5E-01 | 1.1E-02* | 5.7E-01 | 2.4E-01 | 4.1E-05*** | 4.8E-01 |
| **Marketing Abstract** | Coefficient | 4.1458 | 32.9664 | 4.8029 | 14.2334 | -9.2968 | 4.8640 | 10.2999 |
| | P-value | 7.7E-42*** | 3.2E-08*** | 4.6E-01 | 8.5E-03** | 2.1E-01 | 2.7E-01 | 2.7E-02* |
| **Author Abstract** | Coefficient | 6.8639 | 46.7001 | -72.6194 | 28.0646 | 54.9804 | 2.1743 | 0.5490 |
| | P-value | 3.3E-16*** | 1.3E-02* | 4.4E-04*** | 1.0E-01 | 1.9E-02* | 8.8E-01 | 9.7E-01 |

**Supplementary Table 1: Increasing text length of titles, inventors' abstracts, and marketing abstracts.** Linear regression models for studying the increasing length of titles, inventors' abstracts, and marketing abstracts over time (unit: word). Each data point is one invention and the dependent variables (DV) are the word count of titles, inventors' abstracts, and marketing abstracts. We found a significant temporal trend of the increasing length of titles, inventors' abstracts, and marketing abstracts over time. Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Two-tailed tests.

| Category | 1980-2000 | 2000-2005 | 20005-2010 | 2010-2015 | 2015-2020 |
|---|---|---|---|---|---|
| **Biology** | 120 (143) | 141 (227) | 156 (686) | 189 (969) | 234 (895) |
| **Chemistry** | 113 (155) | 141 (231) | 150 (691) | 192 (915) | 233 (813) |
| **Engineering** | 124 (152) | 143 (278) | 150 (736) | 195 (973) | 236 (858) |
| **Materials** | 117 (160) | 141 (251) | 149 (737) | 190 (970) | 233 (865) |
| **Electronics** | 120 (85) | 140 (147) | 150 (323) | 197 (429) | 239 (319) |
| **Energy** | 141 (27) | 129 (63) | 160 (224) | 208 (332) | 256 (256) |

**Supplementary Table 2: Increasing word count of marketing abstracts.** Median word count of marketing abstracts over time across different categories. Numbers in parentheses indicate sample sizes. Inventions with missing marketing abstracts are omitted. Overall, we observed a substantial increase in marketing abstracts' length across all the categories.

| DV: adjective percentage | year | Category: Biology | Category: Chemistry | Category: Engineering | Category: Materials | Category: Electronics | Category: Energy |
|---|---|---|---|---|---|---|---|
| **Coefficient** | 0.0035 | 0.0151 | 0.0038 | 0.0117 | -0.0006 | 0.0069 | 0.0104 |
| **P-value** | 2.4E-52*** | 2.3E-03** | 4.8E-01 | 1.1E-02* | 9.2E-01 | 8.1E-02 | 1.4E-02* |

**Supplementary Table 3: Increasing percentage of adjectives in titles.** Linear regression model for the percentage of adjectives in titles. The dependent variable is the percentage of adjectives in the title (i.e., number of adjectives/title word count). There is a significant increase in the percentage of adjectives in titles over time. Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Two-tailed tests.

| DV: net income rank | | ADJ percentage | Text Length | year | Category: Biology | Category: Chemistry | Category: Engineering | Category: Materials | Category: Electronics | Category: Energy |
|---|---|---|---|---|---|---|---|---|---|---|
| **Marketing Abstract** | Coefficient | -0.6904 | -0.0002 | -0.0011 | -0.0375 | -0.0105 | -0.0068 | -0.0429 | -0.0102 | -0.0347 |
| | P-value | 0.000014*** | 0.000219*** | 0.175074 | 0.014354* | 0.531031 | 0.621378 | 0.027437* | 0.374708 | 0.003691** |
| **Author Abstract** | Coefficient | -0.3959 | -2.711E-05 | 0.0010 | -0.0280 | -0.0145 | -0.0003 | -0.0481 | -0.0126 | -0.0327 |
| | P-value | 0.000027*** | 0.013142* | 0.111065 | 0.037805* | 0.322166 | 0.981662 | 0.003999* | 0.239490 | 0.003684** |

**Supplementary Table 4: Net income rank correlates negatively with abstract length and adjective percentage.** Linear regression models the correlation between adjective percentage, text length, and net income rank. The dependent variable is income rank (net income normalized by year). After controlling for categories, we found that adjective percentage and text length negatively correlated with net income rank on both author and marketing abstracts. This indicates that abstracts of high-income inventions tend to be more concise and use fewer adjectives. Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Two-tailed tests.

| DV: female percentage | year | Category: Biology | Category: Chemistry | Category: Engineering | Category: Materials | Category: Electronics | Category: Energy |
|---|---|---|---|---|---|---|---|
| **Coefficient** | 0.02 | 0.47 | -0.10 | -0.22 | -0.14 | -0.59 | -0.11 |
| **P-value** | 7.9E-07*** | 1.3E-07*** | 2.3E-01 | 5.3E-04*** | 1.4E-01 | 1E-19*** | 8.3E-02* |

**Supplementary Table 5: Increasing female participation.** Logistic regression model for the gender composition of inventors. Each data point is one inventor and the dependent variable is whether the author is female. There is a statistically significant increase in the percentage of female inventors over the years after controlling for invention categories. In addition, there is statistically significantly more female faculty participation in biology and less female inventor participation in engineering and electronics. Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Two-tailed tests.

| DV: inventor team size | year | Category: Biology | Category: Chemistry | Category: Engineering | Category: Materials | Category: Electronics | Category: Energy |
|---|---|---|---|---|---|---|---|
| **Coefficient** | 0.035 | 0.139 | -0.244 | 0.003 | 0.319 | 0.086 | 0.099 |
| **P-value** | 1.7E-30*** | 4.4E-02* | 4.1E-04*** | 9.5E-01 | 1.2E-04*** | 1.3E-01 | 5.8E-02 |

**Supplementary Table 6: Growing number of authors per invention.** Linear regression model for the number of inventors per invention. There is a statistically significant increase in the number of inventors per invention over time, indicating an increasingly collaborative invention environment. Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Two-tailed tests.

| DV: net income rank | First-time Team | year | Category: Biology | Category: Chemistry | Category: Engineering | Category: Materials | Category: Electronics | Category: Energy |
|---|---|---|---|---|---|---|---|---|
| **Coefficient** | 0.0280 | 0.0010 | -0.0354 | -0.0104 | -0.0026 | -0.0487 | -0.0201 | -0.0366 |
| **P-value** | 0.016* | 0.083 | 0.006** | 0.469 | 0.830 | 0.003** | 0.053 | 0.001** |

**Supplementary Table 7: First-time inventors generate higher net income than repeat inventors.** A logistic regression model compares the inventions of first-time inventors and repeat inventors. The dependent variable is income rank (net income normalized by year). The independent variable “first-time team” indicates the team consists of only first-time inventors. After controlling for categories, results suggest that first-time inventors tend to have higher net incomes on their inventions. Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Two-tailed tests.

| DV: fraction of self-licensing | year | Category: Biology | Category: Chemistry | Category: Engineering | Category: Materials | Category: Electronics | Category: Energy |
|---|---|---|---|---|---|---|---|
| **Coefficient** | 0.002 | 0.075 | -0.035 | 0.014 | 0.026 | -0.049 | -0.009 |
| **P-value** | 4E-02* | 3E-05*** | 8.2E-02 | 4.1E-01 | 2.4E-01 | 6.4E-04*** | 5.6E-01 |

**Supplementary Table 8: Self-Licensing is increasingly popular.** A linear regression model for the self-licensing rate of inventions. There is a statistically significant increase in the self-licensing rate over time (p-value 4.0E-02). In addition, the self-licensing rate is higher in the biology category (p-value 3.0E-05) and lower in the electronics category (p-value 6.4E-04). Note: * p < 0.05, ** p < 0.01, *** p < 0.001. Two-tailed tests.

| DV: net income rank | Self-licensing | year | Category: Biology | Category: Chemistry | Category: Engineering | Category: Materials | Category: Electronics | Category: Energy |
|---|---|---|---|---|---|---|---|---|
| **Coefficient** | 0.059 | 0.001 | -0.040 | -0.011 | -0.003 | -0.053 | -0.018 | -0.036 |
| **P-value** | 4.1E-08*** | 2.8E-01 | 1.6E-03** | 4.4E-01 | 7.7E-01 | 9.4E-04*** | 7.1E-02 | 1.2E-03** |

**Supplementary Table 9: Self-Licensing leads to higher net income.** A linear regression model for the outcome (i.e., net income rank) of self-licensing. After controlling for categories, self-licensed inventions have higher net income (p-value: 4.1E-08). In addition, although the self-licensing rate is higher in the biology category (Supplementary Table 8), the self-licensing inventions in biology have statistically significantly lower net income (p-value: 1.6E-03). Note: * p < 0.05, ** p < 0.01, *** p < 0.001. Two-tailed tests.

| DV: net income rank | Inventor team size | Category: Biology | Category: Chemistry | Category: Engineering | Category: Materials | Category: Electronics | Category: Energy |
|---|---|---|---|---|---|---|---|
| **Coefficient** | -0.0236 | -0.0532 | -0.0195 | 0.0371 | -0.1406 | -0.0513 | -0.0506 |
| **P-value** | 0.002** | 0.220 | 0.637 | 0.233 | 0.005** | 0.113 | 0.098 |

**Supplementary Table 10: Team size vs. Self-Licensing Success.** A linear regression model for the outcome (i.e., net income rank) of inventor team size. Only self-licensed inventions are included in this analysis. After controlling for categories, smaller inventor teams are associated with higher net income rank (p-value: 0.002). Note: * p < 0.05, ** p < 0.01, *** p < 0.001. Two-tailed tests.