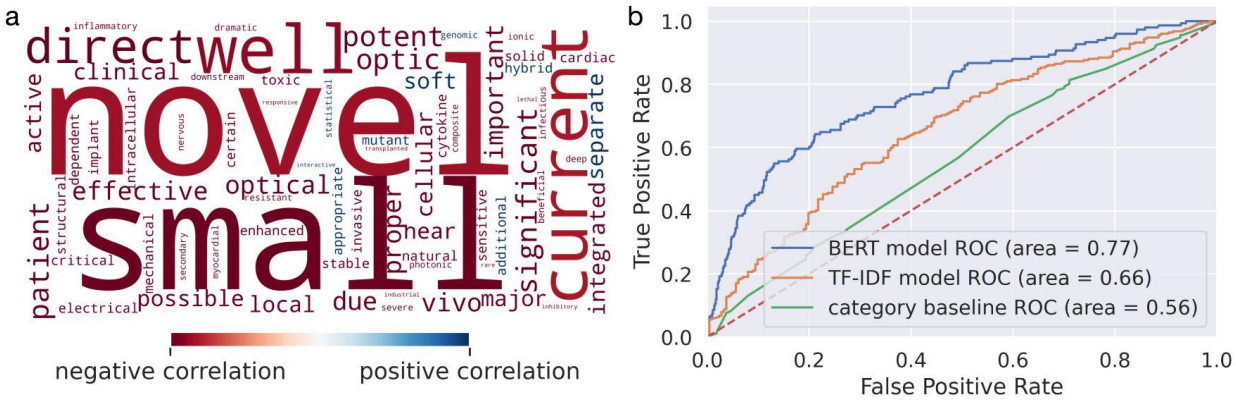
Patterns, Volume *3*

Supplemental information

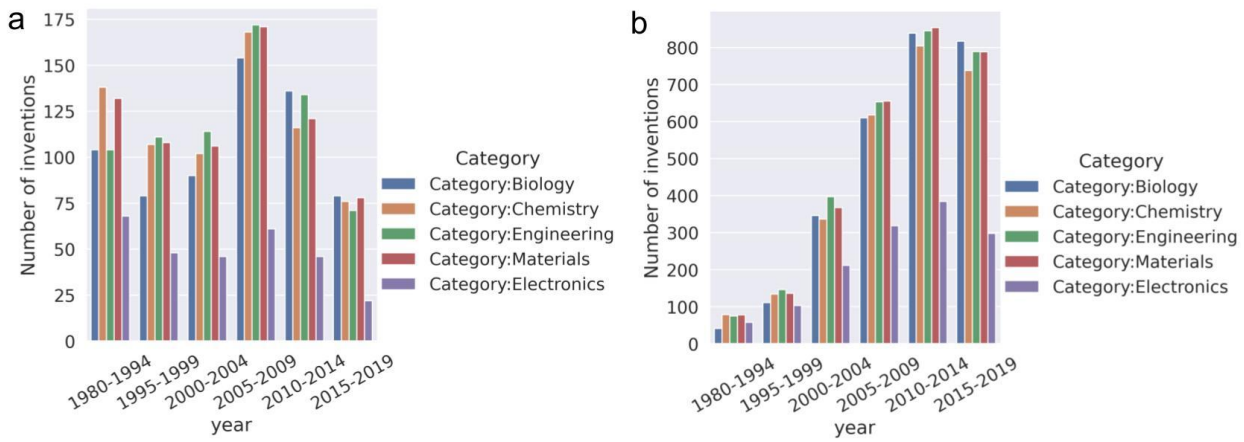
Systematic analysis of 50 years of Stanford University technology transfer and commercialization

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.