

# Legal Tech Start-ups: State of the Art and Trends

Sabrina PRADUROUX <sup>a,1</sup>, Valeria de Paiva <sup>b</sup> and Luigi di Caro <sup>b</sup>

<sup>a</sup> *University of Turin, Italy*

<sup>b</sup> *Nuance Communications, Sunnyvale and University of Turin, Italy*

**Abstract.** We survey the legal tech market, classifying and analyzing a number of legal start-ups, particularly the ones in Silicon Valley, where the first author was based for her graduate summer project with the second author. This kind of survey of the state of the art is inherently incomplete, very tied-up to where and when it is done, and frankly biased towards the interests of the authors. However, if read paying attention to these caveats, this survey can be very useful to practitioners, interested in uncovering the landscapes of the market.

**Keywords.** legal start-ups, state-of-the-art, legal tech, legal informatics market

## 1. Introduction

Legal informatics (legal information technology and its study) is concerned with the social implications of informatics use, as well as with all the applications of informatics in the field of law, such as the storage and the automatic retrieval of sources of law, the automation in law offices and in the judicial administration and all the other uses of the computers in law (data bases; information systems; educational programs; expert systems, computer-aided legal drafting, etc).

Research and development of computational law the branch of legal informatics concerned with the automation and mechanization of legal analysis is growing at fast speed. Many of the new tools and methods come from research in academic institutions, some comes from development done by the big law practices, but most, or so it seems to us, comes from legal start-ups. LEgal start-ups are small companies with one or two founders, a few dedicated hackers and sometimes some venture capital. The statistics are somewhat difficult to verify, but the Stanford CodeX group (<http://tech.law.stanford.edu/>) mentions at the time of writing (October 2016) 576 companies in the space of legal innovation tech companies. Clearly we will not be able to survey all of those. But we aim to indicate in the next section, in broad brushstrokes the main features of this landscape of legal companies.

---

<sup>1</sup>Corresponding Author: Sabrina Praduroux, University of Turin, Italy, E-mail:sabrina.praduroux@unito.it.

## 2. A Landscape of Legal Start-ups

Broadly speaking we are interested in describing what has been called the Legal Tech market. Legal Tech, like its older and more substantial ‘brother’ FinTech (Financial Technology) covers companies (mostly startups) utilising technology to build products solving problems faced both by the legal industry (i.e. law firms, corporates etc.) and consumers of legal services.

Some blogs and analysts<sup>2</sup> have claimed that, while financing activity to startups targeting financial services has spiked over the past five years, startups in the area of legal services have seen no such boom. Since 2011, global legal tech companies have raised just \$739M in aggregate funding, according to the same source. This is despite the ostensible opportunities in the multi-billion dollar legal industry. Of course to measure **growth** of an industry one needs to define what constitutes Legal Tech and boundaries differ depending on the commentator.

In this work, we are not worried about economic growth of the market, quarter by quarter, but instead we are concerned with detecting the most useful and most feasible technologies, aligned with our own academic profiles. The writers have profiles associated with work in Artificial Intelligence (AI), logic and ontologies, as well as Natural Language Processing (NLP) and tools. Thus we spent a long time considering the kinds of classification of companies in Legal Tech that we should help us in our usual academic concerns.

Whether the growth is explosive or not, whether it could or should be faster or not, in any case, we are seeing increasing numbers of well funded legal startups. This in turn raises the profile of the industry and provides further validation for startups looking to get those first customers in a notoriously risk-averse market. The legal services market has historically been viewed as one that is hard to penetrate, when it comes to knowledge acquisition, as liability, therefore costs, can escalate rapidly. The introduction of marketplace models and startups focused on document services is improving this situation. The increased transparency is presenting opportunities for startups to compete with the more well-established players.

Advances in natural language processing have enabled people to build solutions addressing various domains within the overall legal tech market, and we discuss these verticals in the next subsection. Language solutions are taking advantage of large data sets to assist with the automation of certain low level repetitive tasks. The opportunities here to reduce costs are significant as law firms tend to bill by the hour. As this technology improves and the data sets they work on are scaled up it could be possible for solutions to be built to automate more advanced work [3].

We are seeing the emergence of startups that are aiming to go beyond the automation of documents and of repetitive tasks and which aim to provide additional insights into Legal Research. There is an increasing interest in the space from both those within the law and those building solutions from outside the law, aimed at the sector. As a representative example, that has been vastly discussed in the news we can discuss the start-up “DoNotPay”, created by Joshua Browder. Wikipedia describes the launching of this legal start-up thus “On 12 January 2015 it was announced that Browder created the UK’s first ‘robot lawyer’. He ultimately hopes to replace ”25,000 ex-

---

<sup>2</sup>See <https://www.cbinsights.com/blog/legal-tech-startup-funding-2016>.

exploitative lawyers” with robots which can respond to questions with human emotions powered by artificial intelligence.” The fact that this company has saved British motorists in excess of \$4 million (according to Fortune magazine in May 2016, <http://fortune.com/2016/05/21/bots-rise-up/>) shows that this kind of company is worth paying attention to.

We considered a couple of classifications of the kinds of companies in Legal Tech. One classification we looked at was the work of Janine Sickmeyer (<https://lawyerist.com>) which has 12 categories. Some of these could be merged and then they looked more like the classification, coming from the website of CodeX, the Stanford Law School Center for Legal Informatics<sup>3</sup>. When discussing companies, in their database of legal start-ups, the Center came up with a classification<sup>4</sup> that has 8 types of legal companies: Marketplace, Document Automation, Practice Management, Legal Research, Legal Education, Online Dispute Resolution, E-Discovery, and Analytics. Some thinking about these labels and the kinds of tools we are most interested in, gave us the eight categories that we present below, which seem to us not too specific, but not too generic either.

### 3. Types of Legal Tech

We now describe the eight categories of legal tech that we decided to concentrate on. The first three categories are similar to what has happened in other areas of human activity, where computers have been incorporated into the existing workflows to help with data ingesting and management, using spreadsheets, databases and email and online forums to grow business. For each of the categories we give some labels we have seen associated with the category.

1. Lawyer Marketplace– Lawyer-to-Lawyer Outsourcing – Social and Referral Networks. These are online marketplaces connecting lawyers with clients, either end users or other lawyers.
2. Document Automation and Assembly – DIY Legal Forms and Contracts This category encompasses the design of systems and workflows that assist in the creation of electronic documents. These include logic-based systems that use segments of pre-existing text and/or data to assemble a new document.
3. Practice Management – Case Management for Specific Practice Areas – Legal billing. Practice and case management software provides attorneys with convenient methods for effectively managing client and case information, including contacts, calendar and meeting information, documents, and other specifics. All that is involved in facilitating automation in law practices can be considered practice/case management software.

The main features and functions of case management packages are:

- Case Management Information on cases and matters can be made accessible through a centralized database. This database can manage to-do lists; can provide fast and flexible searching; can check conflicts of interest; and can check statutes of limitations, for example.

---

<sup>3</sup><https://law.stanford.edu/codex-the-stanford-center-for-legal-informatics>

<sup>4</sup><http://tech.law.stanford.edu/>

- Time Tracking Records are systems of billable time on an hourly, contingent, transactional, or user defined fee, computed individually or firm-wide. Links to time, billing, and accounting programs are essential parts of these systems. They can also generate client invoices; link to other time tracking and accounting programs; and create reports for individual billing attorneys.
  - Document Assembly and Drafts documents. These require links to word processing programs and templates to facilitate creation of most common documents.
  - Contact Management tracking systems log, and store details about phone calls, e-mails, and other correspondence. They can also provide callback reminders and deadline tracking for processes.
  - Calendaring and docketing software allows staff to view tasks, deadlines, appointments, and meetings by day, week, month, or year. They can calculate calendar dates and schedule appointments and meetings.
4. Legal Research. Legal search engines based on advanced search technology from the fields of artificial intelligence, data mining, and natural language processing, with different characteristics and features are available.
  5. Predictive Analytics and Litigation Data Mining. Predictive analytics is the analysis of data through statistical or mathematical techniques that results in meaningful relationships being identified in the data. These results can then be used for better prediction of future events and better decision-making. Predictive modeling of litigation management provides the information needed at the beginning of a juridical process to improve it.
  6. Electronic discovery (also called e-discovery, ediscovery, eDiscovery, or e-Discovery). This is the electronic aspect of identifying, collecting and producing electronically stored information (ESI) in response to a request for production in a law suit or investigation. ESI includes, but is not limited to, emails, documents, presentations, databases, voicemail, audio and video files, social media, and web sites. This is one hard problem as the law mandates that all legal evidence need to be uncovered in law suits and the enormity of the task is staggering.
  7. Online dispute resolution (ODR). This uses technology, especially the Internet to solve disputes out-of-court through an Alternative Dispute Resolution procedure. There are two basic branches of ODR, both based on different kinds of technology. The first branch may be called Technology based, it refers to those systems where technology plays an active role in conducting the dispute resolution. A prominent example of technology-based ODR systems are blind-bidding systems. The technology uses multivariate algorithms to help parties arrive at the optimal outcome. The second branch of ODR consists of technology-assisted solutions. Technology-assisted ODR refers to the use of technology to augment Alternative Dispute Resolution processes that exist independently of the technology.
  8. Data security technologies. These are intended to protect confidentiality of data that is exchanged in client/server data transfers. Fundamental to these technologies is the use of proven, industry-standard encryption algorithms for data protection.

Our classification is not without problems: for example, we have decided to give e-discovery its own class, when we could have considered it part of practice manage-

ment (as it is part of the workflow of practices) or litigation data mining/analytics or even the more generic legal research. All of litigation data mining could also be considered a subset of legal research. But our classification is pragmatic and since most of innovation occurring in legal research seems to occur in these two categories (e-discovery and litigation data mining), it seems reasonable to give them each its own top category.

Also our last top category can be considered not strictly 'legal' technology, as computer security experts would claim it as their territory. While many lawyers would prefer not to deal the mechanics of ensuring privacy and confidentiality of transactions, this seems a field where engineering by itself is not enough and the the legal profession must interact with computer experts to ensure that the tools developed satisfy their needs. Similarly, there is plenty of work on the borders between Medical Informatics, the Law and AI that need to be addressed. Given that biology is providing data that is far beyond the ability of humans to fully analyze and that the ability to integrate data about patients from streams genomics, personal monitoring and electronic health records to accurately diagnose and choose treatments, it is imperative to think about the anticipated evolution of these capabilities and how they might impact economic, social, political and cultural activities. But the scope of our research has to be made feasible, so we will not dedicate any attention to either Medical Informatics or Data security technologies or even online dispute resolution, because the tools that we have at our disposal are not the most suitable for this kinds of technology.

Another area that is really important nowadays, and one where our tools (Natural Language processing, knowledge bases, logics and ontologies) can help, but we will not try to address in this report is Intellectual Property laws and management. Patents and their multiple issues require specific subject matter expertise that we could not master in the short time we dedicated to the task of surveying the landscape of legal startups. Similarly we will not discuss the area of Legal Education: it requires too much expertise in Law, and it is borderline to our concerns.

#### **4. Industrial Trends**

Altogether e-discovery (# 7) and data management in general (including litigation data mining (# 5), as well as other forms of legal research) seem to be two of the fastest-growing segments of the legal technology market. Data from a report on eDiscovery (Software and Services) Market Trends by Global Industry Analysts, available at [www.strategyr.com/MarketResearch/eDiscovery\\\_Software\\\_and\\\_Service\\\_Market\\\_Trends.asp](http://www.strategyr.com/MarketResearch/eDiscovery\_Software\_and\_Service\_Market\_Trends.asp) informs us that

The global market for e-discovery (Software and Services) is projected to reach US\$11.6 billion by 2020, driven by growing demand from governments and private enterprises, rise in criminal prosecutions and civil litigations, and increased investigational admissibility of digital data.

While the growing trend is easy to see, the numbers on these kinds of predictions are much harder to come by. There are too many so-called specialists online and many do not explain their sources. Nevertheless it is clear that the digital data explosion is creating challenges for all information technology and legal departments in enterprises all over the globe, who need to efficiently handle and manage data for use in active litigation and/or in internal investigations.

## 5. Academic Legal Informatics

Law schools, like large law firms, do not leap into new technological directions easily. But there are some exciting things happening in some of the law schools. Stanford (CodeX), Georgetown (Iron Tech Lawyer), Suffolk (Institute on Law Practice Technology and Innovation) and Chicago-Kent (Center for Access to Justice and Technology) have programs that bring new legal tech into the classroom <http://legal-tech-blog.de/2289-2>.

Using the same categories that we discussed for cataloguing start-up companies we can try to classify the academic projects in Legal Informatics. Clearly schools and the NGOs have the possibility and some even may say the duty of using technology to promote access to justice. This is one of the topical issues of the American Bar Association (ABA) report [2]), and also of the UN Millenium Development Goals (Goal 16 of the Sustainable Development Goals is dedicated to the promotion of peaceful and inclusive societies for sustainable development, the provision of access to justice for all, and building effective, accountable institutions at all levels. <http://www.un.org/sustainabledevelopment/sustainable-development-goals/>). There are many ways to use technology to further access to justice, as described in [1]. The State should invest money for these kind of research projects, and it might be a good sector for private investments as well.

### *Document Automation and Assembly*

Stanford Computable Contracts Initiative (SCCI) (CODEX, Stanford) works on developing a universal Contract Definition Language that will allow terms and conditions to be represented in machine-understandable way. As a result, computers might be able to eventually process and reason over the contracts automatically with a guaranteed degree of accuracy. Fair Document (another CODEX, Stanford project) seeks to drive down the cost of legal services by making the process of performing high-volume transactional legal work more efficient. Fair Document wants to do this by automating the generation of a base set of documents and providing collaboration and workflow tools to perfect the documents. Since automated process are always subject to risk, before a client sees any legal output from Fair Document, it is reviewed by a licensed attorney. Other products already exist in the market to drive efficiency through document automation, but they do not involve lawyers.

The project A2J Author (Center for Access to Justice and Technology - IIT Chicago-Kent College of Law) consists of a software tool that delivers greater access to justice for self-represented litigants by enabling non-technical authors from the courts, clerk's offices, legal services programs, and website editors to rapidly build and implement customer friendly web-based interfaces for document assembly.

### *Litigation Data Mining*

The Smart Prosecution Project (CODEX, Stanford) seeks to combine the latest advancements in data mining and data analytics to apply them to the criminal justice system.

The Computational Linguistics and Effective Legal Drafting (CODEX, Stanford) project focuses on using advances in computational linguistics technology to help

lawyers draft more precise and error-free legal documents like contracts, regulations and statutes. The technology underlying this project includes various natural language processing, machine learning, and data mining techniques. Any public or private sector entity that handles legal documents and is concerned about litigation risk would be a potential user of this solution.

### *Lawyer Marketplace*

Legal.io (CODEX): provides a marketplace infrastructure to legal service enterprises, so that they can coordinate talent, services and transactions. Innovative law firms, bar associations, and legal aid organizations, use Legal.io's collaborative, white-label software platform to reduce overhead, while increasing revenues and engagement with new clients.

The Legal.io technology framework was developed out of research, prototyping and testing on LawGives, the first consumer-facing marketplace platform powered by Legal.io. Legal.io combines AI-based algorithms and the design methods into a ready-to-deploy content-, service-, and panel-management solution, that hopes to be custom-tailored to any law firm or association of legal professionals, large or small.

### *Legal Research*

Not surprisingly many of the academic projects concentrate on Legal Research, where the long term payoff might be more important than immediate profit making capabilities. The Computational Law project (CODEX, Stanford) aims at enabling higher degree of automation to achieve better usability and more efficiency of various tasks involving legal reasoning. It focuses on formalization of governmental regulations and enterprise policies, development of automated reasoning procedures for compliance checking, legal planning and regulatory analysis, and developing user-facing computer systems

Wellsettled (CODEX, Stanford) is a searchable database in several areas of law, such as patents, criminal law, and torts. The project Ravel (CODEX, Stanford) focuses on visualization-based search of legal data. Specifically, Ravel is a search platform for lawyers and law students with a clean, collaborative interface that provides visualization of how legal data (primarily cases) rank and connect to each other. In addition, the platform offers features that enable annotation of case law and collaboration across caseboards. Ravel leverages advances in the fields of network analysis and data visualization to show not just a cases relevance but also how case law evolves and how legal topics cluster.

The LexCraft (Legal Information Institute, Cornell Law School) project aims to record, refine, and promulgate best practices for electronic legal information publication.

The OAI4Courts (Legal Information Institute, Cornell Law School) is a project that promotes the federation of independent legal websites into large, useful "virtual collections" that span boundaries. As a technical project, that is a goal that is within easy reach. Well-understood standards for metadata interoperability, particularly the OAI-PMH standard for metadata harvesting, have been widely used for similar purposes in the digital-library world for several years. They are only slowly making their way into the realm of legal information, though. One of the goals of this project is to accelerate that process. At the same time – like any standard – success will depend crucially on buy-in from legal information creators and from those who publish their work.

Autonomous Intelligent Cyber Entity (AiCE) (CODEX, Stanford). This project explores the commercial and legal aspects and implications of an intelligent cyberagent and its evolution into an autonomous intelligent cyber entity (AiCE, pronounced ice). It evaluates and builds functional and operational schemas for standardizing AiCE with an emphasis on reducing waste of judicial resources, increasing e-commerce transactional certainty and expanding into new frontiers for e-commerce interactivity on business to business (B2B), business to customers (B2C) and customers to customers (C2C) levels.

Designing and Understanding Forensic Bayesian Networks (CODEX, Stanford) with Arguments and Scenarios. Evidence based on statistics can easily lead to errors. This project aims to help prevent this sort of error from occurring. The projects approach is to link the successful statistical modelling technique of Bayesian networks to models that effectively dovetail legal argumentation and scenario construction in the legal world.

Computational law research and development (MIT). Their task is to reformat legislative data, refactor legal code and republish the law as a public digital service. Deeper economic, jurisprudential and socio-political dimensions of the Algorithmic Law project will need to be addressed over longer arcs of time primarily by other stakeholders. In any event, the substantive content of the law can not be significantly adapted for the future until open, structured data standards are adopted as the common containers of law. They provide a roadmap of the current approach of their project. In Phase 1 they aim to establish workable foundations for Computable Law. The idea is to examine existing examples of law as computable data, including thorough legal markup and thorough incorporation by reference of technical standards into law. Then they will extrapolate scenarios demonstrating public law as algorithms that can be defined, debated and decided as part of open, public and democratic processes. Thirdly they plan to define one or more prototypes, test and evaluate specific computational law use cases in relevant business, legal and socio-technical contexts. Following that, in Phase 2 the idea is to iteratively prototype Public Law and Regulation as computational data and algorithmic expressions. Finally in Phase 3 the plan is to postulate and propose dynamic, emergent laws, legal entities, legal instruments, legal relationships, legal transactions and legal systems, in general. It all seems very ambitious and not really developed enough, yet.

## 6. Comparing Landscapes

One of the main differences between the industrial and the academic landscapes of legal tech should be the presence of not for profit projects in the academic sphere. We simply list a few of these projects, noting that some that are called not for profit, seem to be for profit, as well.

### *Non Profit Projects*

**CourtListener** (Free Law Project, a California non-profit public benefit corporation) is a free and open-source repository, search engine, and research platform for analysis of court opinions and other legal documents. The system is a key set of metadata linked to those opinions: a comprehensive database of judges and other persons involved in the operation of the legal system. **MassLegalHelp** hopes to use the Web in creative ways to improve access to justice for low income and disadvantaged people. They are working



to connect, support, and educate advocates and the general public. The content of the website is written by people within the legal services community.

**Caseloop** (founded in 2015) is a free nationwide lawyer to lawyer directory matched with referral tracking software. They aim to simplify the practice, reduce marketing costs, build networks and boost lawyer's income from referral fees. Caseloop allows attorneys to list their firm information in their nationwide directory for free, refer cases to each other, track their progress, and obtain referral fees.

**Arbitrator Intelligence** is a non-profit, interactive informational network that increases and equalizes access to critical information in the arbitrator selection process. Arbitrator Intelligences preliminary start-up phase, organized around a Pilot Project, concluded on January 14, 2015. Arbitrator Intelligence plans to collect quantitative feedback from users and counsel about key features of arbitrator decision making. Information will be collected through surveys allowing users to provide feedback on specific questions such as case management, evidence taking, and award rendering. When (and if) fully developed, Arbitrator Intelligence will allow members to search accumulated information to aid in their arbitrator selection process.

## 7. Conclusions

There is considerable criticism of researchers in Legal Informatics that they do not pay enough attention to the pain points of the legal professionals and instead dedicate their time and public funds to pursuing their philosophical leanings. To counteract this kind of criticism we decided to survey the more applied landscape of legal technical startups in the San Francisco Bay Area and to summarize our efforts, from the perspective of researchers, interested in NLP, logics and ontologies, eager to apply our tools to real life and actionable problems of lawyers and users of the law. While it is clear that in the European Union, a big effort of consolidating laws and jurisprudence in different languages, is a necessary first step, in the United States the multilinguality is not so important. In the more applied side of the law in the US, the automation of documents seems the main goal, followed closely by the automation of the practices workflow.

E-discovery is on the other hand the hard nut to crack and the one where the new tools of AI might be very useful. eDiscovery software and services have become necessary tools for enterprises aiming to resourcefully respond to legal, regulatory and investigational requirements. Aggravated by ever-growing volumes of electronically stored information (ESI), enterprises are facing mounting challenges in collection, review and storing of digital data for use in litigation, regulatory and investigation processes. Massive increases in electronic business-to-business (B2B) and business-to-consumer (B2C) communication, growing use of smartphones, tablets and other Internet-enabled systems for enterprise communication, and rising importance of big data in day-to-day business operations are driving ESI volumes in enterprise environments. The data explosion is creating challenges for IT and legal departments in all companies to efficiently handle and manage data for use in active litigation and/or internal investigations. Enterprises of all sizes, including large companies, multinational companies, and small and medium-sized businesses (SMBs) are feeling the need for eDiscovery software and services. Additionally, the growing number of criminal investigations and civil litigations embroiling enterprise operations, along with the need for internal investigations for compliance and

regulatory purposes, will put more pressure on enterprises to prioritize deployment of eDiscovery software and services.

## **References**

- [1] James E. Cabral, Abhijeet Chavan, Bonnie Rose Hough Linda Rexer Jane Ribadeneyra Thomas M. Clarke, John Greacen, and Richard Zorza. Using technology to enhance access to justice. *Harvard Journal of Law and Technology*, 26 (1)(243):29–pages, 2012.
- [2] Commission on the future of legal services. *Report on the Future of LEGAL SERVICES in the United States*. ABA, 2016.
- [3] Hughes-Jehan Vibert, Pierre Jouvelot, and Benoît Pin. Legivoc – Connecting Law in a Changing World. *Journal of Open Access to Law*, 1(1):19–pages, 2013.