Nice (France), January 4th 2017

A few thoughts by the Cloem team in response to the article *"I Think, Therefore I Invent: Creative Computers and the Future of Patent Law"* by Professor Ryan Abbott (University of Surrey School of Law, University of California, Los Angeles, David Geffen School of Medicine), published in the Boston College Law Review (Vol. 57, No. 4, 2016), quoting Cloem on page 1118.

January 4th 2017

Artificial Intelligence. The article by Professor Abbott considers patent laws in the advent of Artificial Intelligence (A.I.). Cloem focuses on a specialized branch of A.I.: Natural Language Processing (NLP), and more specifically on linguistic manipulations of patent documents.

Cloem vision. Cloem in many ways can be seen as a "hack" of the patent system. As inventions are expressed in natural language, Natural Language Processing technologies can be advantageously applied to patents. The application of NLP is somehow more limited than A.I., but not less challenging and rich of consequences. A vision of ours is to "merge" copyright and patent laws. We try to "exhaust" all possibilities of expressions of an idea, with actual textual renderings (in analogical, written out form, i.e. immediately accessible to the human cognition). It is possible to publish patent claims and variants thereof, as patent documents are generally exempted from copyright (i.e. there is no authorization requirement to create derivative rights).

Combinatorial space. The statistical analysis of the entire patent corpus indicates that the word count of the independent claim 1 is between 120 and 150 words. We observe that the length of claim 1 is finite (and rather low) while the vocabulary is also limited (clarity criteria reduce the combinatorial space). Considering a claim of 150 words, each being selected among 2000 ones, elementary mathematics shows that there are 2000x2000x...x2000 (150 times) thus 2000^150 possible sentences most of which are of course ungrammatical. After considering *grammatical* sentences, along specific sentence structure rules (e.g. preamble/characterizing part), the presence of semantically "empty" words or invariant parts, the combinatorial space is considerably reduced. Cloem can currently manipulate 30 slots with up to 200 possible fillers for each slot. Brutalizing the permutation space is therefore manageable. (- Incidentally, it is striking to consider that most if not all human inventions can be expressed in so few words.)

Incentives to build creative machines. About "*incentives*" to build creative machines, as developed by Professor Abbott: we think these arguments and perspectives are interesting, but we do envision some radical extremes. In the case of an effective protection, the private use of A.I. is a significant risk (winner-takes-all, preemption of IP). If no protection can be obtained, the public domain may well increase at unprecedented scales (quantity and density), for example due to "global scorched earth policies", i.e. massive defensive publishing. With or without protection, an extended use of secrets is possible (as it is increasingly the case today, where non-discoverable inventions are buried deep in data centers or in hardware circuits). In other words, the outcomes probably can remain very diverse (as of today), as a consequence of a complex and subtle system of incentives. Cloem technology might equally lead to "public domain" operators, or to A.I. trolls.

Game changers. The high cost of the current patent submission process generally favors the concentration of patent rights and oligarchic structures (e.g. IBM), lone inventors generally playing only a marginal role. A.I. may accelerate this process, because building an A.I. system is capital-intensive (for now). As seen with the last patent reform, the oligarchy of tech companies is rather happy with the status quo and may be willing to fight to protect the current system. Who has the incentive to actually move things ahead? Tech companies may not be the natural candidates to change the system. More agile companies can still create a specialized A.I., for example a startup such as Cloem.

Liability. The article by Professor Abbott considers "*consumer protection*" (page 1079). As explained in our Terms of Service, we do not "test" cloem texts, i.e. we have *a priori* no (actual i.e. real) reduction to practice. The vast majority of inventions filed today by humans are not tested either. Some can embed hidden perpetual motions, imperfections, insufficiencies or the like. For this reason and others, we discourage as much as possible any reduction to practice of a cloem text, and we disclaim any liability in this matter. As a future additional layer to our model, Cloem is developing (conceptual/constructive) reduction to practice (by embedding common sense such as specialized and general ontologies, automated/semantic reasoning, machine learning the usefulness of previous patents, etc.) A related aspect of this topic is that with IT inventions in particular, the skills of the skilled person are steadily increasing and Common General Knowledge can appear to be a "black hole" (evidence is hard to bring on the table by Examiners, due to the complexity of the corpus embodying his skills). We decided to embody the "skilled person" in our model, concretely. We generally agree that an - essential - question lies in the liability of an "AI" system. To our knowledge, there are few answers available (see the developments related to automated cars)

Inventing model. Professor Abbott argues page 1080 that it is a "*matter of time before computers are responsible for most innovation*". We respectfully disagree. It is relatively "easy" to generate, hard to generate efficiently and it is very hard to select or evaluate. In our view, inventing is two-fold process, and both steps are equally important. The term "filter" page 1085 appears to be key and this direction of research appears wide open.

Inventorship. Professor Abbott argues page 1081 that "*computer inventorship would incentivize creative machines, then new scientific advances*". Our team again is not so assertive. Dissemination of ideas will remain key among human societies. The "social" diffusion of inventions is not likely to be automatable (it has been hardly predictable for centuries). Computational inventions might also literally "flood" inventors and scientists. Human attention *in fine* remains limited. Professor Abbott also recalls page 1082 the rationale of copyright to prevent corporate ownership and seems to argue that computer ownership determining inventorship would fit well US patent laws. We observe that computer ownership is not always clear: difference between hardware and software, use of the Cloud, virtual machines, dynamic mashups, etc. The infrastructure also may be increasingly concentrated (see Jaron Lanier and GAFA data centers

Man-machine interactions. The article suggests that machines may "refocus" or otherwise "assist" human inventors. Co-creation is what we are also envisioning, and in fact what we propose via our website. Cloem tries to let inventors focus on their intuitions, while we get the grammar right. In some ways, we're not trying to replace the inventor, we are rather trying to partly replace the routine work of a patent attorney. A patent professional – *routinely* – add, delete or replace words in patent claims. We do scale, automate, fasten and optimize these steps, by linguistic

manipulations. At Cloem, we do envision that intellectual routine work can be outsourced, freeing valuable human attention time. At Cloem, we try to alleviate the pain of patent attorneys and to unleash intuitions of inventors. For example, drawings are generally a waste of time (except for understandability and possibly litigation). It is significantly better to describe as much as possible and to multiply embodiments (same words with different orders in a same paragraph, leading to better searches, etc.)

Based on mathematics. Professor Abbott underlines the Human Authorship Requirement page 1100 *"... operates randomly or automatically without any creative input or intervention from a human author..."* About this point, we try to have as little randomness as possible in our model and to involve human users as much as possible. In Cloem's model and program, we have tons of embedded choices in the model (e.g. structure of the program, of the sentence, of the parts we select in specifications, in vocabulary lists, thresholds we apply, etc.). Writing software truly is authorship. In the end, the creation of Cloems is not random but based on mathematics and human expertise.

Weak signals. A particular challenge, part of this new interaction model, is that Cloem is striving to solve is to detect weak signals (in the patent corpus and more generally in the scientific literature). In the foreseeable future, Cloem will allow detecting and then "amplifying" such weak signals. For example, after a new word is detected early e.g. "touch" in relation with "screens", it can be propagated to the entire corpus (e.g. medical devices). As a corollary, speed matters. Cloem intends to develop schemes similar to the ones existing in High Frequency Trading.

NLP changes the game. Professor Abbott discusses "mental act". For us, this is obsolete Law, or Case Law. Prehistoric considerations. A business method claim equals an IT invention modulo a few words. From an NLP perspective, statutory rejections are completely obsolete. We think that linguistic manipulations can entirely renew the way we deal with inventions.

Private use of A.I. Professor Abbott discusses that the A.I origin or contribution to an invention can be hidden to patent offices by applicants. We fully agree and suggest that the private use of A.I. is a very general question (in trading floors, banks, etc.) *Automatonophobia* is a reality. More general aspects related to this subject might comprise algorithmic governance, transparency and the like. We add that it is very difficult, and increasingly so with advances in technology and linguistic knowledge, to distinguish between a CG (computer-generated) text and a human text (if not impossible for a short text).

Recognition or discovery of A.I. inventions. Professor Abbott develops page 1098 the idea of "discovery" or "recognition" of computational inventions by human beings. We note that in the novel of Borges, some people go insane in the gold rush in the Library of Babel. Should we direct our scientists and inventors in such labyrinths? The man-machine interaction model is yet to be invented.

Thinking in silos. Versus the thinking in silos of patent offices, the coexistence of more and more inventions is positive for the collection of taxes by patent offices, but not very fair for inventors, who in particular use general search engines, and are totally agnostic of "technical domains". Patent laws basically ignore that Google searches exist, juxtaposing results. There are no archives or logs of the displayed results. One of our current goals at Cloem is to create massive transpositions between scientific domains (e.g. from biology to IT and vice versa), i.e. artificial cross-pollination between scientific fields. For example irrigation principles by the Egyptian

hydraulics can find some echoes in today's microfluidics. Oil drilling principles can be applied to medical devices, and vice-versa.

Quantified skilled person. Professor Abbott discusses the "skilled person". From the very beginning of Cloem, we envisioned a quantified skilled person. We ingested the entirety of Wikipedia, tons of scientific manuals, TRIZ methodology, etc. In many ways, the Common General Knowledge remains a "black hole" and is difficult to model (it equals the general model). Ingesting Wikipedia for example does not directly translate into words or rules. One could consider that variants created for a claim ("cloems") represent "embodiments" of this Common General Knowledge.

Legal support requirement becomes excessive. It would be possible to consider that mechanic extensions (cloem texts) are due (can be assigned) to the initial human inventor. We think "legal support" (e.g. written description, Article 123 in Europe) according to patent laws are (or tend to be) unfair for inventors. A (clear) intuition should allow getting a patent, i.e. to cover the (main) variants of the seminal idea. When an inventor provides an invention, he "implicitly" discloses a "class" of inventions. It is fair to let the machine explicitly fill in the gaps, relying on the encoding of the patent attorney's skills. Nowadays, improvement patents are too easily obtained and end up in mutual dependencies. Patent drafting unavoidable defects turn against the initial applicant. At Cloem, we would like to "secure" rights with some close neighboring tolerance zones, the latter being quantitatively defined. In our model, the seminal idea by a human can be "exhausted" by the computer and this can largely prevent any (direct) improvement patent thereon (modulo a textual distance). The European problem-solution approach is *in fine* arbitrary (e.g. dependent on the initial upstream search, which is necessarily incomplete, etc.) and such an approach shall eventually be replaced by quantitative methods. Patent quality can essentially be reduced to a smart word count. The advent of quants in IP is unavoidable.

About Cloem. Cloem has developed algorithmic patenting for years, with advanced NLP technologies. We now have generated more than 100 billion cloem texts ("claim+poem=cloem"), based on human users' inputs (more than 1 million claim trees), and counting... The choice of vocabulary to create variants is human or machine based.