* Indian community faces a “Digital Divide” due to dominance of English as mode of communication in higher education, judiciary, corporate sector and Public administration at Central level whereas the government in states work in their respective regional languages [1]
* India has 22 scheduled languages. While 99% of the population speak one of these scheduled languages in various dialects (which number in the thousands)[2], according to Census 2011, the total percentage of English speakers is at 10%, and that too is skewed towards the urban population.[3] Hence, there lies a need for developing NLP architectures for facilitating flow of digital content and information in and between local, national and international levels.
* The above also means that a large percentage of the literate population is monolingual or bilingual, and across 22 languages, an accessible, easy-to-use and intuitively developed system is required, which enables intercommunication.
* While traditionally NLP has been approached with statistical methods such as Hidden Markov Machines (HMM), Support vector machine(SVM), Conditional Random Field(CRF), Naive Bayes(NB), etc, which take a large amount of tagged/annotated data (corpus) to statistically analyze and learn the language characteristics[4], the research into deep learning or ‘connectionist approach’[4] has gained impetus due to (i) the simplicity of the solution in rapidly prototyping and establishing practically effective systems (ii) the lower cost of annotation of the training data[5], and the fact that they attempt to more closely emulate the learning process of biological brains, among other reasons. [4][6][7]
* Particularly, the collection of a uniform corpus and standard datasets for training models remains a challenge across all regional languages. The large number of morphological variations across Indic languages also contributes to this issue.[8][9]
* Most of the Indian population accesses digital content through smartphones. In 2019, the number of smartphone users in the country passed 500 million[10], and is estimated to increase to 850 million by 2022[11]. This, hence, also makes smartphones and smartphone apps in particular an ideal platform on which to launch NLP applications for the wider population, and directly help facilitate flow of information past language barriers.

1. ( Kurian, C. and Kannan Balakrishnan, K., 2008. Natural Language Processing in India Prospects and Challenges. In Proceedings of the International Conference on “Recent Trends in Computational Science.)
2. <https://indianexpress.com/article/india/more-than-19500-mother-tongues-spoken-in-india-census-5241056/>
3. <https://www.livemint.com/news/india/in-india-who-speaks-in-english-and-where-1557814101428.html>
4. <https://arxiv.org/ftp/arxiv/papers/2102/2102.00214.pdf>
5. <https://arxiv.org/ftp/arxiv/papers/1907/1907.12437.pdf>
6. <https://arxiv.org/pdf/1610.09565.pdf>
7. <https://www.aclweb.org/anthology/W09-0438.pdf>
8. <https://arxiv.org/pdf/2005.00085.pdf>
9. <https://towardsdatascience.com/nlp-for-indian-languages-310d1d8a10b6>
10. <https://www.news18.com/news/tech/smartphone-users-in-india-crossed-500-million-in-2019-states-report-2479529.html>
11. <https://telecom.economictimes.indiatimes.com/news/indian-to-have-820-million-smartphone-users-by-2022/76876183>