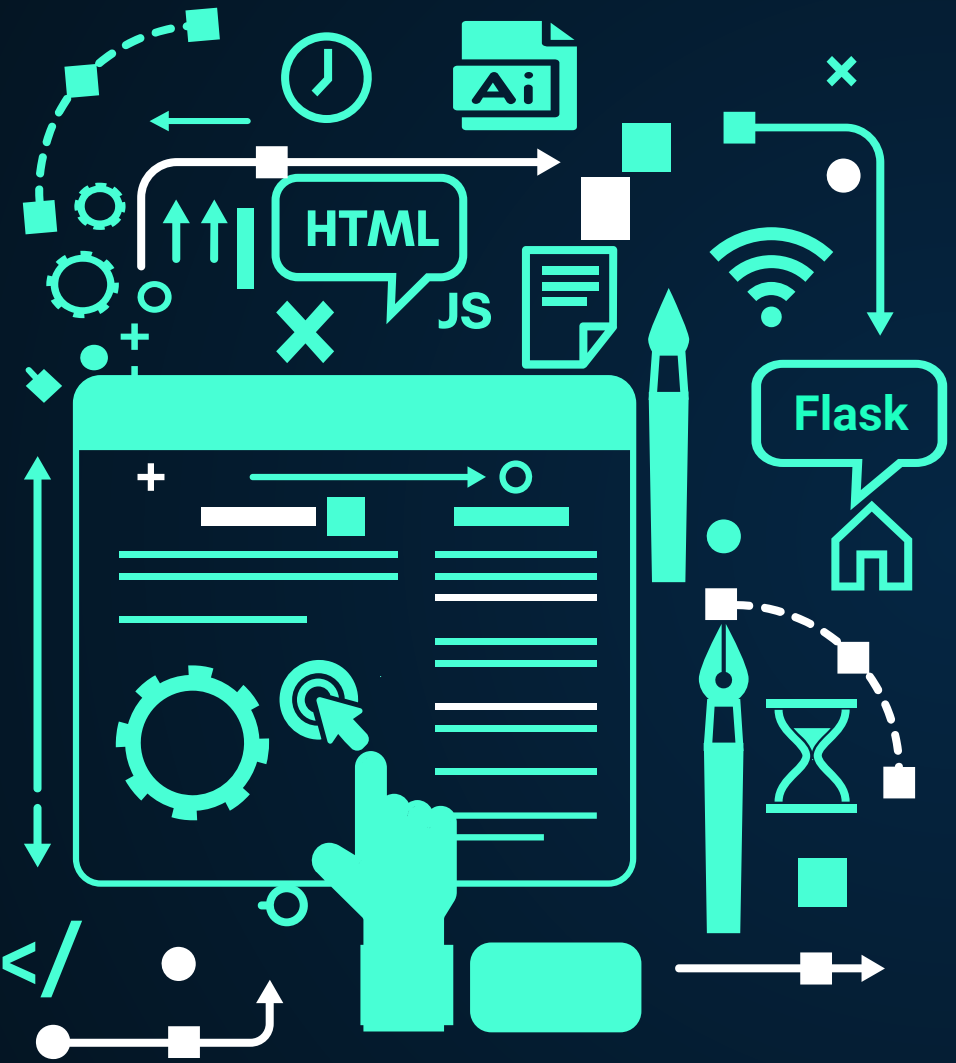


TLDR TEXT SUMMARIZER -NS1230

VSCube_SIH2022

Problem statement by:
Policy Planning Cell, Ministry of Information and Broadcasting

WRITER'S BLOCK

A boring test

KENNETH D. MAHERN, University of Denver, Colorado, U.S.

Do you write boring papers? Of course not. Your papers are interesting. Right? Not right? Well, maybe once in a while a manuscript escapes you that is boring, but usually it's certainly not boring. Right? How could your papers be boring? Your work's not boring, so your writing must not be boring. You read and reread your manuscripts, and each is interesting, concise, valuable, maybe even elegant. You don't put them to bed easily. In addition, your colleagues always say, "Interesting paper!" So, tell me, if your writing isn't boring, what is the font of boring papers? Other people, you say.

Consider other people-not me who write boring papers. I've looked at it. I've asked and been given the same answers other people. Confused, I searched for the "other people" and have yet to find any. This means the density of "other people" is sparse and seemingly below the threshold for producing the existing density of boring papers. Hence, the other-people explanation is suspect.

So, we come back to same question: Who produces the boring papers? After further research, the only answers are: (a) a new property of the ether, that undetectable, all pervasive "stuff" that 19th century scientists thought filled space, used to explain the confusing behavior of electromagnetism; or (b) there are those among us who are not aware of the condition of their papers! We can safely eliminate the former—just think of the authorship and copyright problems it would cause. Would you trust a paper written by undetectable stuff? Unfortunately, that leaves only the latter.

Are you wondering whether you are a member of that club? Perhaps it would help if you took a membership or qualification test. Funny thing, I just happen to have a boring test. Here's a list of questions. Answer them all on the same page. What makes something boring? Eric Haseltine writes ("The beauty of boredom" in the March 2000 issue of *Discover*): "The answer lies buried deep in our nerve cells, which automatically dampen their initial excited response to stimuli every subsequent time a stimulus occurs. Those neurons also enhance response to things that change—especially that which changes quickly. We probably evolved this way because our ancestors got bored with the same old thing and went from attending to what moved in a tree (such as a panther) than to the tree itself. Boredom, as a reaction to a static environment, turns down the level of neural excitation so that a new stimulus (such as that panther) stands out more. Boredom is nature's way of extinguishing a perch light to see the fireflies."

Haseltine seems to be saying that system drive readers to boredom by being monotonous, predictable, repetitious, circuitous, or long-winded. Well, I'm not sure that I buy into that test. You're safe—there's no score. Judge your own answers and draw your own conclusion(s). I have parenthetically included some points to ponder when considering your answers.

Your general philosophy. Do you believe that your work must be as interesting to all as it is to you? (This falls under the umbrella of "writer's folly"—i.e., my work is supremely important and will make everyone do things differently in the future. Reality check! Unless your name is Newton or Einstein and you've found a cure for cancer, world hunger, global warming, or a comparable issue, chances are your work is, like that of most, a small but valuable piece of humanity's knowledge.)

Do you recognize that persuasion is the heart of successful technical writing? Do you write to allow the physiologist, the scientist, the engineer, to develop, attract, and persuade their readers. Readers become bored and move to something else. Successful writers recognize that they are salespersons of their work and present solid arguments for its value, utility, and benefit. Do you write persuasively?

Do you stop writing because you are done or because you're out of interest, time, or steam? (If writing about this material bores you, what do readers find?)

Your abstract. Do you compose a concise, stand-alone, and terse summary of the new information in your paper or do you just slap something together? (A boring abstract clues the reader that the rest of the paper won't be much better.)

Do you read your abstract hated labor? (If so, readers will know this and assume, again, that the rest of your writing is the same.)

Does your abstract droll on? (Unless your article is a tome, your abstract should be about 6–8 exacting, information-packed paragraphs.)

Do you stubbornly cling to the belief that abstracts must be in passive voice? (Passive voice easily lulls prey to long, indirect sentences that say very little, while appearing to say a lot. Typically, abstracts in passive voice are a boring information void.)

Your introduction. Do you copy your abstract and use it as the first paragraph in the introduction? (Boring!)

Do you feel the success of an introduction is based on the number of references cited, the more the better? Or, said a bit differently, do you feel that your work is a high-sentence description of work by other people make

"TOO LONG; DIDN'T READ."

While the thirst for consumption of information has gone up, the attention span of readers has also gone down to a large extent! Often long pieces of texts go unnoticed despite their importance and not every citizen is always invested in reading long press releases!
Often most critical information goes unnoticed.

BUT WE FOUND A SOLUTION!

The Daily Telegraph

SURRENDER PLAN IN KENYA FAILS

MAU MAU MAY HAVE KILLED NEGOTIATORS

'WE HIT HARDER' SAYS GEN. ERSKINE

ROCKET-FIRING JETS TO BE USED—OFFENSIVE STARTS

BUNS FOR ALL RATINGS IN NEW CARRIERS

COUNCILS MUST KEEP UP CIVIL DEFENCE

Minister's Letter to Coventry

H-BOMB VIEW CHALLENGED

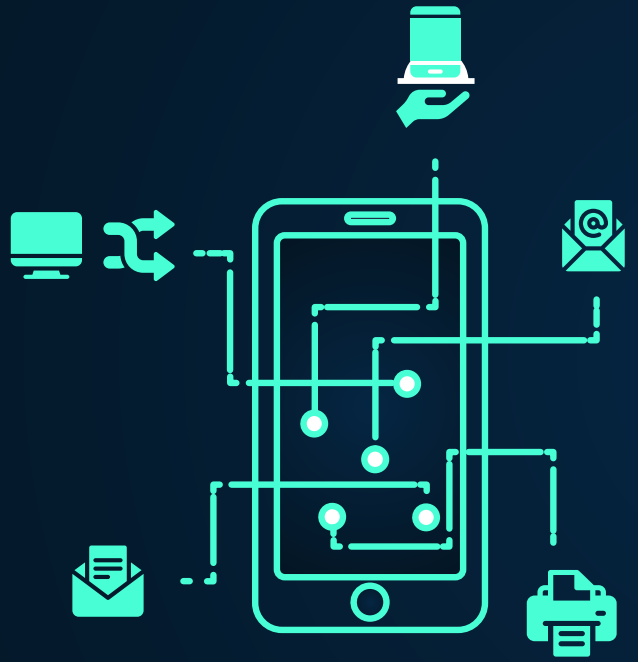
MR. DULLES IN LONDON FOR INDO-CHINA TALKS

"DANGER" IN THE FAR EAST

BLAZING CAR LEAP BY MIKE HAWTHORN

U.S. AIR FERRY TO INDO-CHINA

LATE NEWS

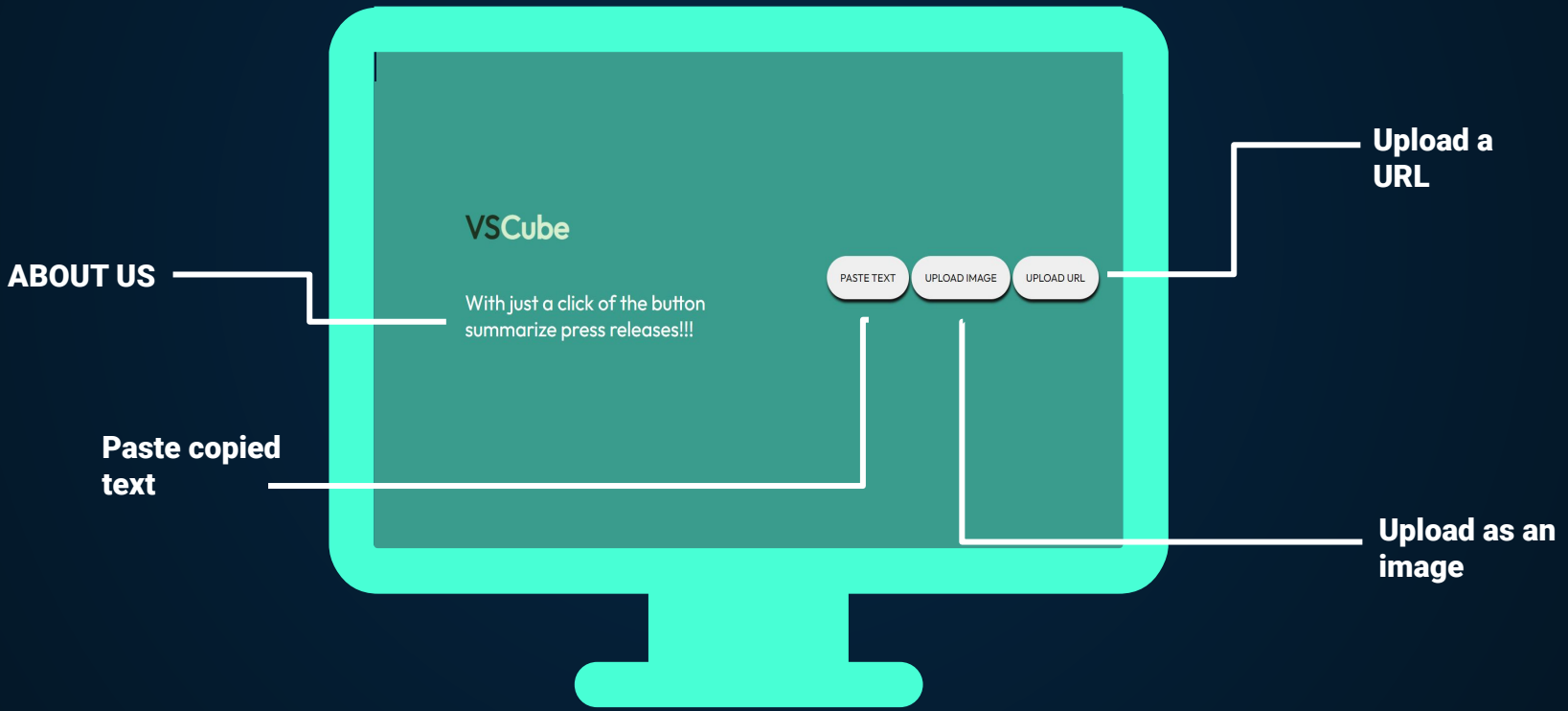


THE PROJECT

Thus to fill this gap, we require short summaries that can grab the attention of the reader! Converting this to visually appealing animations and infographics will also help further! Which is why, we, team VSCube_SIH2022 chose to work on this topic!

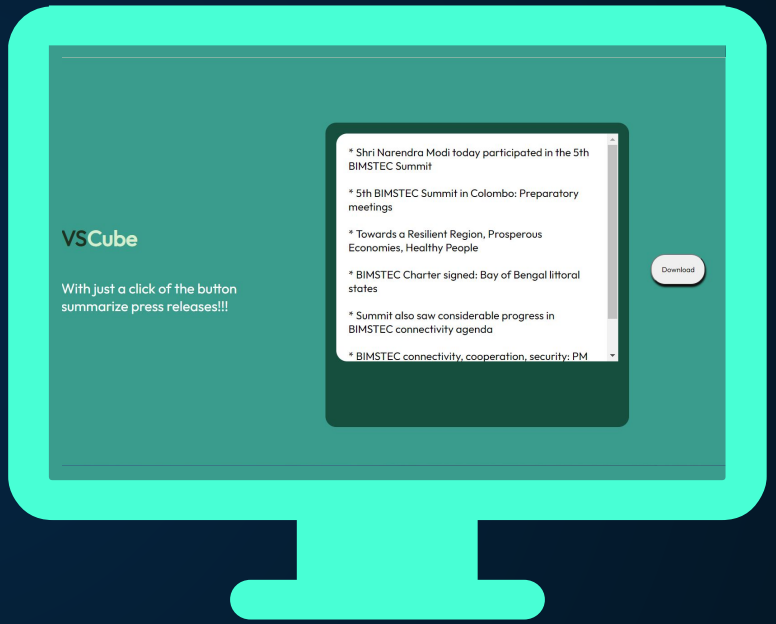
Our version of this project aims to use the existing NLP - python module, T5, to summarise the text document

WORKING PROTOTYPE OF OUR WEBSITE



SUMMARISE ANY PRESS RELEASE IN A CLICK!

SUBMIT THE TEXT



SUMMARY GENERATED!

OUR WORKFLOW

DATA SET ACQUISITION

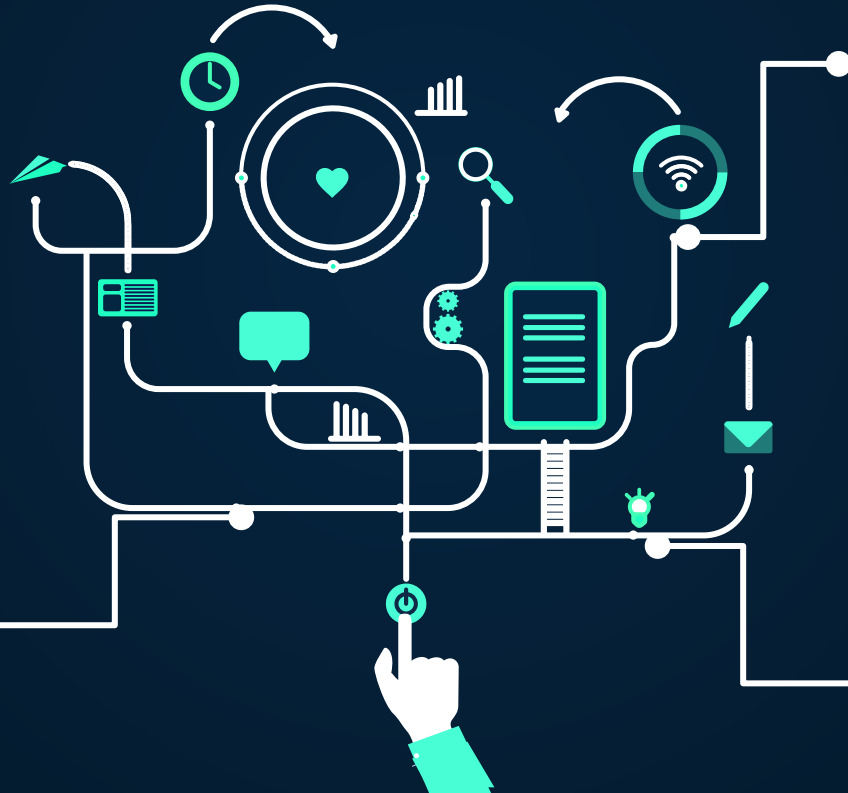
ML MODEL EXPLORATION

WORKING ON EFFICIENCY OF SUMMARIZATION

WEB PAGE INTEGRATION - UI

INCORPORATING FUTURE SCOPE

UNIQUE DATASET

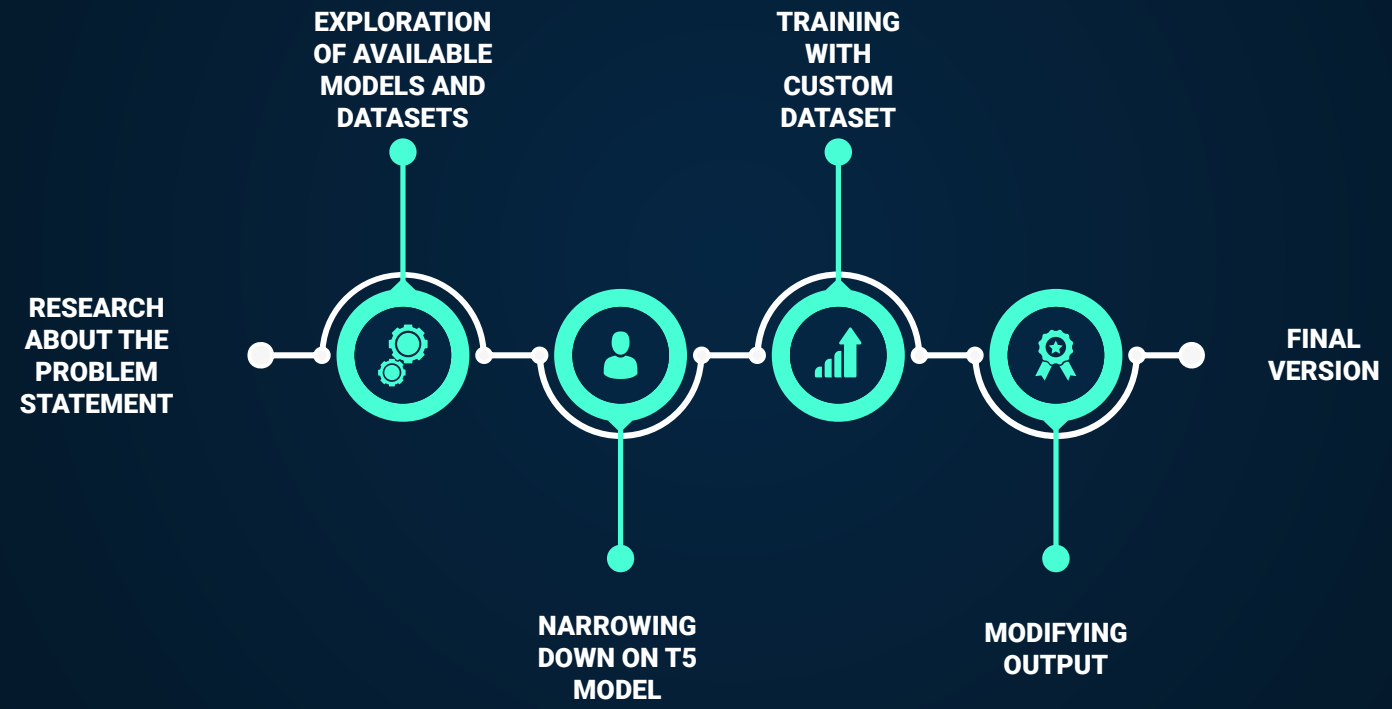


ABSTRACTIVE SUMMARIES GENERATED FROM EXISTING SUMMARIZERS USED FOR TRAINING

WEB SCRAPING ***THE OFFICIAL PIB WEBSITE*** FOR MORE RELEVANT DATASET

SEVERAL MINISTRIES AND THEIR PRESS RELEASES CONSIDERED

ML ALGORITHM BUILDING



OUR USP

SEVERAL MODES OF INPUT AVAILABLE

Users can use upload **images** and **URLs** in addition to keying in texts.

CUSTOMIZED SUMMARIZER

A customised summarizer to exclusively suit the needs of **Policy Planning Cell**

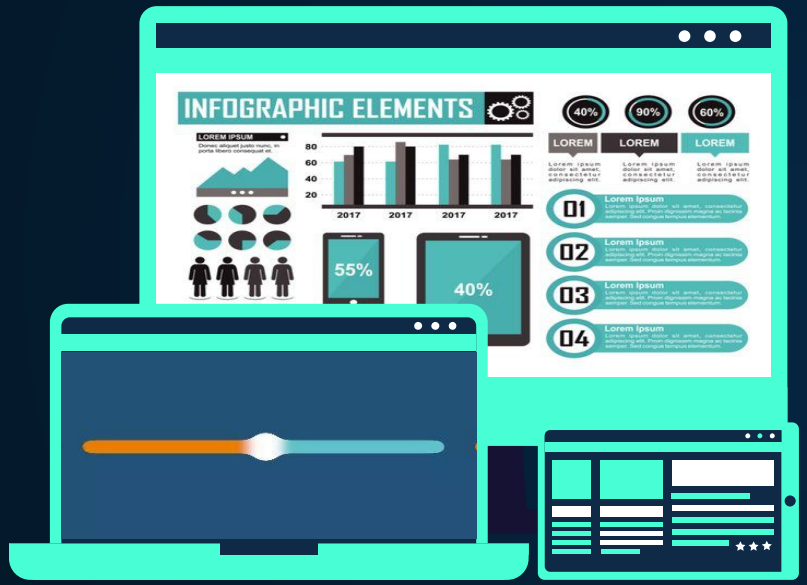
SUMMARY AVAILABLE IN JPEG format.

DATASET CREATION

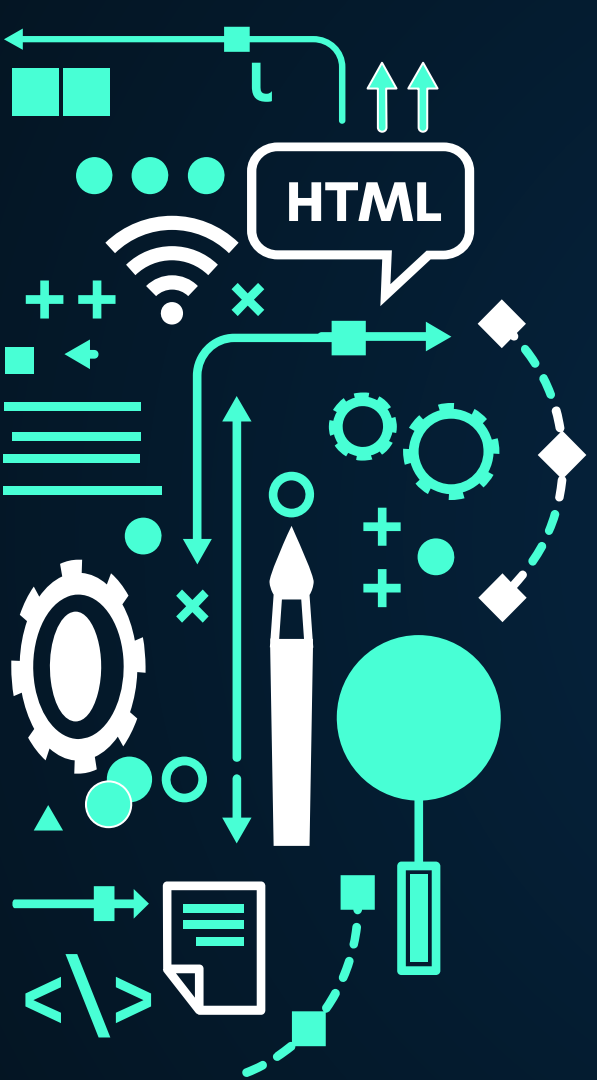
Unique dataset personalised for ***PRESS RELEASE*** summarization

This makes the **MODEL MORE EFFICIENT IN SUMMARIZING**

WHAT WE PLAN TO DO



- INFOGRAPHICS - DASHBOARD
- KEYWORD EXTRACTION
- READER ADAPTABILITY



THANK YOU!

TEAM: VSCube_SIH2022

Team Leader: Varsha Yeshwanth
Team Members:
Subhashree M
Vishal V
K Sowmya
Vignesh SM
Sowmya Ganesan

II Year CSE Students
Sri Sivasubramaniya Nadar College of Engineering, Chennai