

Human Computer Interaction (Hindi mein)

Professor Rajiv Ratn Shah

**Department of Computer Science and Engineering
Institute IIT Madras**

AI-Powered Tools for Content Generation and Analysis: Firefly, Audino

Lec39

[Sangeet] Hey everyone mera naam Rohan Kumar hai. Main ek research assistant hoon Dr. Rajiv Ratan Shah ke under. Mera primary kaam audio AI, NLP aur web development ke around rehta hai. Toh aaj main aapko kuch popular AI domains ke example dikhane wala hoon. Jaise text to image, text to video and audio processing. Toh chaliye shuru karte hain. Sabse pehle hum dekhte hain ek hum text pass karte hain. Descriptive type of text humein dena hota hai. Uske basis pe model ek image generate karta hai. Seedhi taur se baat karein agar toh model ek prompt leta hai. Prompt ke basis pe generate image generate karta hai. Prompt kya hota hai? Prompt ek semantic meaning hota hai ek text ka. Jo bhi text hum daalte hain uska semantic meaning pehle generate kiya jaata hai. Phir us semantic meaning ke tehat image ko generate kiya jaata hai. Toh woh hote hain hamare text to image. Phir aate hain hamare paas text to video. Same concept pe kaam karte hain. Text to image ke concept pe kaam karte hain. Yahan pe image hi generate hota hai but sequence of images generate hote hain. Bahut saare images generate hote hain. Phir un images ko combine karke ek video ka roop de diya jaata hai. Toh usse hum text to video convert karte hain. Third hamara topic rehne wala hai audio processing. Like audio jo hota hai jo bhi hum audio upload karte hain ya audio dete hain. Audio se sabse pehle hum extract karte hain jo text hote hain audios mein bole gaye vaky jo hote hain woh hum extract karte hain. Phir unki madad se hum un data ka kuch bhi kar sakte hain. Like summarization, chapterization, translation kuch bhi hum kar sakte hain. Use hum kehte hain audio processing. Toh chaliye hum start karte hain text to image tool se. Toh aaj ke time pe net pe bahut saare tools available hain. Aaj main aapko dikhane wala hoon Adobe Firefly. Adobe Firefly free toh nahi hai but kuch usage like 10 images aap generate kar sakte hain free free access mein. Toh chaliye dekhte hain ki text to image kaam kaise karta hai. Toh yahan pe hamare paas Adobe Firefly hai. Ab hum dekhte hain ki yeh kaise kaam karta hai? Ismein text to image kaam kaise karta hai? Maine yahan pe ek prompt daal rakha hai jise hum descriptive type question kahenge. Yahan pe aap padh sakte hain cute brown puppy sitting on a green grass. Toh hum keh rahe hain ek puppy hona chahiye jo ki green grass pe baitha hona chahiye. Under a bright blue sky. Saaf aasman neele aasman ke neeche hona chahiye. There are few white clouds in the sky. Kuch safed baadal bhi hone chahiye sky mein. And colorful flower around the puppy. Bahut saare colorful flowers hone chahiye puppy ke aaspaas. Toh chaliye hum ise generate karte hain aur dekhte hain. Ab model kya karega? Is text ko lega aur iske embeddings generate karega. Jaise-jaise model embedding generate karega us

embeddings ko pass kiya jayega model pe aur woh image generate kare karega. Hum dekh sakte hain hamare paas output mein chaar images aa chuke hain aur jo ki saare text jo bhi maine likha hai usko follow kar rahe hain. Har ek part yahan pe bhi ek puppies hai. Saare mein puppies hain. Sky jo hai sky jo hai woh blue hai. Bahut saare safed baadal bhi hain. Phool bhi aaspaas aapko dikhai de rahe honge. Toh jo model hota hai woh model poori tarah se follow karta hai. Jo bhi hum text daal rahe hote hain. Main aapko yeh link de dunga. Aap is pe jaiye aur ismein click karke dekhiye. Yahan pe dog ki jagah aap cat daal sakte hain. Change kar sakte hain aur dekh sakte hain model kaise output de raha hai. Hamare paas kuch yahan pe edit features bhi hain jo ki hum effect waghera daal sakte hain. Main aapke upar chhod raha hoon. Aap ise play karein aur iske saath familiar ho jaayein. Chaliye ab baat karte hain text to video tools ki. Toh text to video tools mein bhi bahut saare net par available hain. Like Pictory Lumen5 in video. Aaj hum dekhenge ki in video kaise kaam karta hai. Jaise woh text leta hai aur video humein return mein deta hai. Toh in video bhi free tool nahi hai but kuch samay ke liye aap ise use kar sakte hain. Ek do video aap ise generate kar sakte hain free mein. Aaiye dekhiye dekhte hain kaise yeh generate karta hai. Toh yeh raha hamara in video. Humne same prompt daal rakha hai jo ki humne pehle daal rakha tha text to image mein cute brown puppy wala. Yeh wala humne same text daal rakha hai video AI mein aur hum yahan pe video generate kiye hain. Video generate karne mein time lagta hai isliye maine pehle se hi generate kara rakha hai. Agar main is video ko play karoon toh main aapko dikha sakta hoon ki yahan pe bhi ek puppy hai. Ismein aawaz mere ko pata nahi aapko sunai de raha hai ya nahi. Yahan pe bhi ek puppy hai jo ki aaspaas jo hain flowers hain. Bahut saare colorful flowers toh nahi but hain. Flowers hain aur yeh video jo hai jo bhi humne text prompt daala hai usko yeh cater karta hai. Use woh follow karta hai. Kahin pe bhi yeh apne mann se kuch cheezein nahi daal raha hai. Toh yeh hamara humne dekha kaise text to video kaam karta hai. Humne prompt same daala tha. Main aapke upar chhod raha hoon. Aap ise khelein tweak karein aur apni apne hisaab se generate karein. Thank you. Hamara third audio processing. Audio processing mein bahut hi popular open source tool hai. Uska naam hai Audino. Audino kya karta hai? Audio video dono mein hi specialized hai. Aur sabse khaas baat hai yeh Indian languages ke upar trained hai. Toh chaliye aaiye check karte hain kaise Audino kaam karta hai. Kya-kya iske features hain. Toh hamare paas hai yeh Audino tool. Audino ke paas Audino ke chaar pillars hain. Pehla hai Content AI, doosra hai Meeting AI, teesra hai Call Recording AI, chautha hai Hate Speech AI. Content AI mein basically jo bhi hum upload karenge uska jo bhi audio ya video upload karenge uska humein summarization milega, chapterization milega aur hum us video us audio se interact kar sakte hain, baatein kar sakte hain, chat kar sakte hain. Second aata hai hamara Meeting AI. Meeting AI mein hum kya karte hain? Jab bhi hum upload karte hain, koi bhi meeting upload karte hain, toh humein unmein se insights milte hain ki meeting ka kya agenda hai? Meeting ka kya topic hai? Kya conclusion nikla hai? Similarly call recording AI and hate speech AI main ek-ek karke aapko samjhata hoon. Pehle hum andar chal ke dekhte hain content AI kaisa dikhta hai. Toh content AI mein yahan pe maine ek video pehle se hi upload kar rakha hai. Yeh video bhi ho sakta hai, audio bhi ho sakta hai. Kisi bhi form of audio data hona chahiye. Hum yahan par maine ek YouTube video upload kiya hai jo

ki ek Samsung ke phone ka features bata raha hai. Yahan pe hum dekh sakte hain ki yahan pe mere paas pehle woh detail hai jo bhi upload kiye gaye hain. Theek hai? Phir hamare paas yahan pe ek summary generate ho ke aa jaata hai. Sabse pehle model kya karta hai? Jo bhi audio hote hain audio ka transcription kiya jaata hai. Transcription hote hain audio ke jo textual form hote hain usmein convert kiya jaata hai. Usko phir jaisa ki maine pehle bataya tha aapko embeddings generate kiye jaate hain. Us embeddings ke baad iska ek summary generate kiya jaata hai. Toh yeh ek summary hai. Humein poori audio dekhne ki zaroorat nahi hai. Poori video ki zaroorat nahi hai. AI ki help se hum ek summarization nikaal sakte hain aur woh hum dekh sakte hain. Hamare paas yahan pe feature milta hai English, Hindi, Arabic. Kisi teen language mein hum convert kar sakte hain usko real time mein. Phir hamare paas aa jaate hain chapters. Chapters yeh hain jaisa ki humne audio ya video upload kiya hai. Har ek block mein kya important points bole gaye hain woh yahan pe mention hai. Taaki hum yahan pe directly click karke directly hum yahan se sun paayein ki kya bola jaa raha hai, kya topic hai. Ek tarah se indexes hote hain jaise pe books ke starting mein. Usi tarah ke is audio ke indexes yahan pe generate kiye gaye hain. After this hum is video ya audio se chat bhi kar sakte hain. Jaisa ki yahan pe dekh sakte hain ki humne chat kiya hai which device is the video talking about? Toh isne Audino ne answer kiya hai. This video is talking about the Samsung Galaxy device. Phir hum yahan dekh bhi sakte hain. Yahan poocha gaya hai ki is it M20 model? So yeh keh raha hai ki iske paas information nahi hai. Toh yeh jo bhi audio mein information honge woh usi basis pe answer karta hai. Toh yahan pe humne dekha ki kaise content AI help karta hai audios ko summarize karne mein, chapterize karne mein aur audio ke discussion mein. Chaliye aaiye dekhte hain kaise meeting AI kaam karta hai. Meeting AI bhi same functionality pe kaam karta hai. Humein sabse pehle ek audio ya video upload karna hota hai. Yahan pe maine ek meeting video upload kar rakha hai pehle se hi. Ab yahan pe jab maine upload kar rakha jab maine upload kar rakha hai toh AI ke dwara yahan pe ek summary milta hai jismein humein pehle bataya jaa raha hai ki meeting ka agenda kya hai? Yeh video kaafi lamba hai. Video 30 minutes ke approx hai. Agar main aapko dikhana chahoon toh 30 minutes ke approx video hai. 33 minutes aur yeh video ko mere ko poora shuru sunne ki zaroorat nahi hai. AI ki madad se hum yahan se hi directly agenda padh sakte hain within a few minutes. Yahan pe hum dekh sakte hain kya key decisions liye gaye hain is meeting ke tehat. Kya-kya action items rahe hain? Kya next steps hain? Kya meeting sentiments kaise rahe hain? Aur meeting ke intent kya rahe hain? Toh yeh saare detail humein bas sirf aur sirf audio upload karna hai aur yeh saare information humein mil jayenge. Hum similar way mein hum isse chat bhi kar sakte hain. Is is audio ke sandarbh mein hum isse baat bhi kar sakte hain. Toh yahan pe hamara second pillar aata hai yahan pe hamara second pillar aata hai meeting AI. Ab third hum dekhte hain call recording AI. Hum sab jaante hain ki jab bhi hum kisi bhi call recording hum sab jaante hain ki jab hum kisi customer care ko call karte hain toh hamare issues ko resolve kiya jaata hai. But yeh bahut hi important hota hai jahan pe bhi hum call kar rahe hain unke side se bhi yeh verify kiya jaaye ki jo agent hain jo calling agent hai unhone resolve kiya hai ya nahi. Toh pehle ki time mein kya hota tha ki log sun ke verify karte the ki haan jo customer ne kaha aur jo agent ne reply kiya woh sahi hai ya nahi kuch galat toh nahi hua toh pehle sun ke kiya

jaata tha ab humein nahi kuch karna humein AI mein upload karna hai aur jaise hi hum upload karte hain humein ek summary mil jaata hai ki client kya ask kiya hai client ne kya poocha hai kya actions rahe hain agent ke kya solution diya gaya hai kya action item hai in future kya complaint hai? Complaint type hai yeh call ya kis type ke calls hain? Intent kya tha calling ka? Toh yahan pe ab humein simply upload karna hai audios ya video aur humein saare details yahan pe mil jayenge. Hum yahan pe similar way mein chaar chat bhi kar sakte hain. Jaisa ki yahan pe hum chat karke yeh pata kar sakte hain. What was the client asking? Client kya maang kya pooch raha hai. Toh, hum dekh sakte hain client was asking about \$33 charge in July 2020? Toh \$33 ka use overcharge ya jis tarah ke bhi charge hue honge woh agent ne resolve kiya. Toh yahan hamara feature ho gaya call recording AI. Last hai hamara hate speech AI. Chaliye dekhte hain hate speech kya karta hai. Hate speech AI bhi similar tarike se kaam karta hai. Humein ek audio aur video upload karna hota hai aur hum yahan pe dekh sakte hain ki jo bhi audio ya video hota hai us video ke jo chunks hote hain har ek chunks ka yeh descriptive detail deta hai. Jaise ki hum dekh sakte hain ki pehla chunk isne capture kiya hai ki 0 minute 0 second se 1 minute 58 second mein controversiality hai. Like jo jo bhi discussion hue na wahan pe thoda controversiality hai. Kya context hai aapko yahin se pata chal raha hai. Uske baad beech mein kuch nahi hua hai. Kuch bhi hate speech detect nahi hua hai. Phir 17 minute 45 second se humein hate speech detect hua hai jo ki controversiality hai. Theek hai? Is tarah se bahut saare Audino AI ke paas hain. Like bahut saare alag-alag categories hain. Like drug abuse, illegal activities, violence, child abuse, etc. Hum yahan par dekh sakte hain ki bahut saare chunks yahan par paaye gaye hain. Like in chunk 18 chunk 8 mein insensitivity bhi paaye gaye paaya gaya hai category. Toh hamare paas 13-14 type ke categories hain jo ki yeh capture karta hai. Similar way mein hum ise chat bhi kar sakte hain. Toh yeh hamara raha fourth chaaron pillar aur yeh raha Audino. Thank you.

[Sangeet]