Human Computer Interaction (Hindi mein)

Professor Rajiv Ratn Shah

Department of Computer Science and Engineering
Institute IIT Madras

Privacy, Security, and HCI: Lecture 11

Lec40

[Sangeet] Namaskar saptah 11 mein aapka punah swagat hai Human Computer Interaction ke is course mein jo ki is baar hum Hindi bhasha mein preshit kar rahe hain. Is saptah mein hum gopniyata, suraksha, HCI ke paripeksh mein charcha karenge. Hum dekhenge kaise hum jab bhi koi maanav kendrit system banate hain toh kaise usmein gopniyata aur security ka dhyan dena bahut zaroori hai. Anyatha aapke jo user honge unke liye pareshaniyon ka saamna karna pad sakta hai. Unko financial loss ho sakta hai. Unko mental stress ho sakta hai. Anxiety ho sakti hai. Agar kuch unki gopniya jaankari bahar aa jaaye toh kaise hum apne application ke madhyam se, apne system ke madhyam se in baaton ka dhyan rakhenge ki kaise hum apne upyogkartaon ki suraksha ko aur unki gopniyata ko dhyan mein rakhte hue ek maanav kendrit system bana sakte hain. Toh chaliye aaiye is baare mein hum yeh saptah charcha karte hain. Iske pehle agar hum ek sankshipt punravritti karein pichle saptah mein humne baat ki thi kaise bada bhasha model LLM aur AI kaise HCI ke saath agar milkar ek system ka nirman karte hain toh kaise woh apne upyogkartaon ke zarooraton ko bade aasani se bade achhe tarike se solve kar sakte hain. Humne baat ki thi kaise bhasha model aur bade bhasha model ek doosre ko ek doosre ke baad aage badhte hain. Kaise language model kis-kis tarike ki samasyaon ko solve karte hain aur jin samasyaon ko language model dhang se solve nahi kar paata hai toh hum bade language model ke madhyam se emergence ke baare mein humne baat ki thi. Jaise humne baat ki thi jaise emergence ki quality hoti hai ki jab hum koi quantitative changes karte hain toh uske corresponding ek bada qualitative changes bhi humein dikhta hai. Jaisa ki hum bade bhasha model mein humne dekha tha jab hum bhasha model se aage badhte hain. Toh uske baad humne bade bhasha model ke architecture ke baare mein bhi baat ki thi. Humne baat ki thi ki kaise alag-alag tarike ke LLMs, multimodal LLMs jo hamare sangyan mein hai jo abhi hum use kar rahe hain. Kaise woh transformer architecture par aadhaarit hai. Attention is all you need waale paper ke baare mein humne aapse charcha ki thi aur humne bataya tha ki is paper ko definitely aapko padhna chahiye. Kaise kaise attention ek aham bhumika nibhata hai transformer architecture mein aur kaise usne poore machine learning paradigm ko change karke rakh diya. Uske baad humne multimodal LLM aur jo LLM hai usse samvaad karne ke liye alag-alag prompting techniques ke baare mein bhi charcha ki thi. Uske baad humne alag-alag applications jo ki HCI aur LLM ko use karte hue ek achhe system ka nirman kar sakte hain us pe baat ki thi. Humne baat ki thi kaise adaptive learning system aap bana sakte hain. Humne baat ki thi kaise

personalize content HCI aur LLM ke madhyam se aap use karke alag-alag applications mein use kar sakte hain. Humne baat ki thi kaise aap LLM ka use karke accessibility aur social inclusion ko bhi apne system mein incorporate kar sakte hain. Uske baad humne ek case study ke madhyam se dikhaya tha kaise ek futuristic mobile phone agent aapke kisi bhi bade jatil se karya ko jo ki anyatha alag-alag application ke madhyam se alag-alag chhote-chhote karya kar poorn kiya jaata hai. Ek step mein yeh kar deta hai aur jisse ki jo upyogkarta hai usko bahut zyada sahuliyat hoti hai apne karya ko karne mein aur kam intervention ke saath kam dikkato ke saath woh bahut aasani se jo bhi uske liye best solution hota hai uska nirdharan karta hai aur last mein humne tutorial aur assignment LLM aur HCI ke upar saptah mein kiye the. Toh chaliye hum is saptah mein aage badhte hain. Yeh is saptah ki rooprekha hai. Hum baat karenge harms kya hai? Deep fake kya hai? Hum baat karenge ki HCI ke perspective mein gopniyata aur security kya hai? Aur hum kaise in baaton ka dhyan rakhenge. Hum ChatGPT ke madhyam se aapko ek demo bhi dikhayenge. Kaise aapke upyogkartaon ki gopniyata suraksha khatre mein pad sakti hai aur kaise un cheezon ko humein dhyan dena chahiye ki yeh jo ek tarah se sensitive information hai woh kisi tarike se leak na ho. Woh kisi tarike se kisi aur ke paas na jaaye aur finally assignment ke madhyam se aap mulyankan apna kar sakte hain ki aapne is adhyay mein kya-kya seekha aur usi ke mutabik aap prashnon ke uttar de sakte hain. Toh agar hum alag-alag in articles ko dekhein toh hum dekh sakte hain kaise alag-alag researchers ne ek tarah se highlight karne ki koshish ki hai. AI ki vajah se alag-alag bade-bade nuksaan bhi ho rahe hain. Jaise ki yahan pe dikhaya gaya hai AI could cause harm if misused by the medical worker. Toh is article mein yahan pe detail mein ek tarah se is pe charcha ki gayi hai ki iske kya-kya nuksaan ho sakte hain? Agar medical worker jo hai usko galat tarike se use kare. Galat tarike se kaise use kar sakte hain? Yeh aap alag-alag points ke madhyam se soch sakte hain. Jaise agar woh kisi tarike se pata kar lete hain ki aapko yeh samasya hai. Usko misuse karke kisi aise log ko bata sakte hain jinko yeh nahi batana chahiye. Aur bhi kai tarike ho sakte hain. Usi tarike se is article mein agar aap dekhein toh bombshell Stanford study find ChatGPT and Google Bard answer medical question with racist debunked theories that harm black patient. Toh ek tarah se is tarah ka pakshpaat poorn karya karna bilkul galat hai. Toh us tarike se humne is is article mein dekha ki kaise ChatGPT aur Google Bard medical questions ke upar racist answer dete hain. CEO behind ChatGPT warns Congress AI could cause harm to the world. Ek tarah se yeh bhi dekh sakte hain ki kaise Sam Altman advocated for number of AI regulations including a new government agency charged with creating government standard for the field. Toh ek tarah se yeh dheron aise article bhare pade hain internet ki duniya mein jahan pe logon ne bataya hai ki agar AI aur machine learning galat tarike se use hoga toh iska bahut hi bada harjana aapke upyogkartaon ko is samaj aur maashre ko bhugatna padega. Toh toh hum kya kar sakte hain? Agar toh sabse pehle toh humein pata karne ki zaroorat hai ki kis-kis tarike ki samasyayein ya kis-kis tarike ka nuksaan hum is tarah ke system aapke upyogkarta ko kar sakte hain, pradaan kar sakte hain. Aur ek baat aur main yahan pe highlight karne ki koshish karunga. Yeh Geoffrey Hinton jinko bilkul kisi parichay ki zaroorat nahi hai. Jinko 2018 mein Turing Award mila tha inke contribution AI mein AI ke mein karne ke liye aur Nobel Prize mila tha physics mein 2024 mein. Toh aise bahut kam

log hote hain jinko aise bade-bade award ek se zyada baar milte hain. Geoffrey Hinton ko ek tarah se yeh award do baar mila. Iske pehle agar aap dekhein toh toh yeh Google ke Google Brain ko ek tarah se head karte the aur inka jo time tha Google Brain mein aur University of Toronto mein as an emeritus professor ke taur par tha. Toh pichle kuch saalon mein humne khaastaur par dekha hai Google chhodne ke baad kyunki Google mein kaam karne ke dauran bhi unhone kaafi aisi cheezein dekhi jiske baad un woh bolne lage in baaton ko ke upar logon ke dhyan ko aakarshit karne lage ki AI agar hum bahut zyada strong banane lagenge toh uska agar misuse hone lagega toh usse bahut hi badi samaj ko aur jo human kind hai unko bahut zyada dikkato ka saamna karna pad sakta hai. Woh ek tarike se kisi aise world ka nirman kar sakte hain jo ki hamare liye bahut hi nuksaandayak hoga. Toh jaisa ki hum dekh sakte hain Geoffrey Hinton ne bola tha I aur aisa unhone Google resign karne ke baad bola aur unhone yeh baat boli ki Google mein rehte hue yeh bolna sambhav nahi tha kyunki again woh company hai company ki policy hai aur Google ke zyada tar product abhi AI based hai. Toh wahan pe unke liye bolna anuchit tha. Toh isi vajah se unhone Google mein jo Google Brain mein jo unka position tha usko unhone chhoda aur uske baad as a as a basically AI ki chintaon ko aage batane ke liye unhone ek tarah se us us group of people ko join kiya jo ki batane lage ki AI ko ek samajhdari purvak regulation ke saath use karna padega. Anyatha ismein bahut hi zyada dikkato ka saamna logon ko karna pad sakta hai. Jaisa ki unhone bola tha. I now think that the idea of super intelligent AI jo ki ek tarah se aadmi se bhi manushyon se bhi zyada samajhdar hai. Could be harmful is something we really should worry about. Kyunki agar woh kal ko apne se decision lene lage toh hamare behalf pe woh financial fraud kar sakta hai. Hamare behalf pe woh alag-alag scandal kar sakta hai. Hamare behalf pe alag-alag crime kar sakta hai. Aur lekin uski jawabdehi kiski hogi? Yeh humein dekhne ki zaroorat hai. Toh the fact that we are creating system that might soon be more intelligent than the human is something that could have very serious consequences. Kyunki jis tarike se jis tarah se bade bhasha model ban rahe hain, jis tarah se unki capabilities hain alag-alag modalities ko samajhne ki chahe woh text ho, chahe image ho, chahe video, audio kisi bhi tarike ki jo modality ya alag-alag information hai, uska woh itne achhi tarike se aur itne kam samay mein samajh pa rahe hain. Human level tak pahunch rahe hain. Agar woh un human se zyada achhe tarike se usko karne lage toh kyunki unke paas ek tarah se itni compute power hai, memory hai jisse ki woh jo manushyon ko kaam kisi kaam ko karne mein kai din lagte the, kai varsh lagte the, woh chand minton ya seconds mein kar lenge aur kaafi had tak bade-bade nuksaan human kind ko kar sakte hain. Toh, again iske dher saare serious consequences ho sakte hain. Isliye yeh bahut zaroori hai ki hum alag-alag hum alag-alag standard ko banayein, guidelines ko batayein aur un guidelines ke madhyam se ensure karein ki hamesha manushya ka jo hit hai woh sarvopari hona chahiye. Kisi bhi keemat pe kisi manushya ko harm nahi pahunchna chahiye. Uske alawa aap unke jeevan ko saral banane ke liye unke karyon ko karne ke liye aap jo kuch bhi karna chahte hain, woh kariye. But kisi bhi keemat par jo aapke manushya hain, jo society hai, jo mankind hai, usko kisi bhi tarike ka koi harm nahi pahunchna chahiye. Toh yeh sab baaton ka dhyan rakhte hue mukhyata main yeh kehna chahta hoon ki aap dekhiye AI vardaan hai ya abhishaap hai yeh is par nirbhar karta hai ki woh kis

tarike se use ho raha hai. Kyunki agar hum dekhein toh yeh pehli baar aisa yeh discussion nahi aaya ki yeh jo cheez banayi gayi hai, yeh jo technology banayi gayi hai, yeh achhi hai ya buri. Kyunki iske pehle agar aap kahenge jab atomic bomb banaya gaya tha, uranium ko use kiya gaya tha toh usko aapke upar hai ki usko aap as a nuclear reactor use karke bijli paida kariye ya poore sheher ko poore desh ko uda dijiye. Aap revolver, machine gun is tarah ki cheezein banate hain toh agar aap usko crime control karne ke liye apni suraksha ke liye border par use karein tab toh theek hai. But agar wahi agar koi aatankwadi galat tarike se use kare logon ko maare toh woh galat ho jaata hai. Toh is wajah se humein yeh pata karne ki zaroorat hai ki jo bhi takneeki hum bana rahe hain usko kis tarike se use karna chahiye. Usko kahan pe use karna chahiye. Kiske dwara use karna chahiye. Agar hum in baaton ka dhyan dein aur toh mujhe lag raha hai ki AI ek bahut hi achha vardaan saabit hoga aur aaj ke time pe yeh kehna galat nahi hoga ki artificial intelligence ek tarah se usi aavishkar ki tarah hai jaise jab electricity ka aavishkar kiya gaya tha aur usne poore industry se leke maanav jeevan ko ek achhi disha mein aur upar ki taraf le gaya toh yeh hamare upar hai ki hum isko kaise achhe se aage use kar sakte hain. Toh chaliye aage badhte hain. Kis-kis tarike ki samasyayein ho sakti hain. Kyunki jab tak hum samasyaon ko nahi janenge tab tak hum uska nidan nahi kar sakte. Toh chaliye sabse pehle jaante hain kya dikkato ka saamna karna pad sakta hai. Chahe woh LLM ho chahe woh artificial intelligence ho. Jab hum usse usko use karke koi model ya system banate hain toh hum dekh sakte hain ki woh bias aur discrimination se grasit ho sakta hai. Ho sakta hai ki usmein koi purvagraha ho. Us purvagraha ke anusaar agar woh koi karya karega toh ek tarah se yeh galat hoga. Agar woh discriminate kisi bhi tarike ka pakshpaat karega toh woh bhi galat hoga. Toh AI system humne dekha hai kaafi time jaisa ki humne news article mein bhi dikhaya tha woh pakshpaat poorn karya kar sakta hai ya woh ek tarah se kisi purvagraha se grasit hoke kisi galat karya ko kar sakta hai toh AI system can unintentionally perpetuate or amplify existing biases present in the data they are trained on kyunki agar humne jo kyunki kaafi baar hamare paas ek tarah se koi control nahi hota hai ki jo bhi data hum ek tarah se machine learning model ko AI model ko train karne ke liye le rahe hain aur agar woh data hi kisi bhi tarike se purvagraha se ya kisi bhi tarike ke discrimination se grasit hai usmein agar woh samasya hai toh jo hum model banayenge woh bhi usi tarike ka banega aur woh bhi usi tarike ka purvagraha aur discrimination karega. So this can lead to discrimination outcome in areas like hiring, lending, policing or health care. Kisi bhi area mein jahan par aap AI machine learning ka use kar rahe hain. Agar jis data pe aapne train kiya hai toh woh galat decision lega. Woh galat matlab ho sakta hai ki hiring recruitment jo karna hai woh galat kare ya woh ek tarah se pakshpaatpoorn tarike se kare. Jaisa ki humne pichle kuch saptah mein yeh bhi discuss kiya tha. Amazon ne ek recruitment tool, hiring tool banaya tha AI based aur baad mein yeh paaya gaya ki woh mahilaon ke prati pakshpaat poorn nirnay leta hai. Mahilaon ko kam select karta hai. Jabki purushon ko zyada select karta hai. Toh ek tarah se is tarah se agar kisi purvagraha ke saath koi model karya karta hai aur yeh karya usi tarike se karta hai jis tarah ke data ke upar yeh train hai toh humein usko pehle sahi karne ki zaroorat hai. Toh example ke taur pe ek aap yeh bhi dekh sakte hain facial recognition system that performs worse on certain ethnic group due to bias training data. Toh aisa humne kai example aur training news article mein bhi

dekha hai ki jo black people hote hain kabhi-kabhi facial recognition system un pe sahi kaam nahi karta aur unko kisi janwar ki tarike se categorize karta hai. Toh yeh bilkul galat hai aur aur aisa isliye hota hai kyunki ho sakta hai ki jo data jis pe train ho woh woh in tarah ke logon ko na include kiya gaya ho ya kam is tarah ka data ho ya galat tarah ka even annotation ya labeling kiya gaya ho. Hum yeh bhi dekhte hain kisi ek prakar ke community ko, kisi ek prakar ke region ko, kisi ek prakar ke dharm ko kisi tarike se label kiya jaata hai. Agar hamara jo data jis pe train hai us tarike ke bias se grasit hai toh phir us tarah ka usi tarike ka karya woh karega jab hum real life mein usko data denge. Toh humein yeh dhyan dene ki zaroorat hai jis data pe hamara kritrim buddhimatta base model train hua hai woh kisi bhi tarike ke purvagraha ya pakshpaat poorn nirnayon se ya us tarah ke data se press na ho. Toh chaliye aur usi tarike se agla jo dikkat ho sakta hai gopniyata, aap gopniyata ka ullanghan ho sakta hai. Toh jaise AI system often rely on large amount of personal data. Misuse or improper handling of this data can lead to privacy breaches. Jaisa ki humne kai dheron aise examples mein dekha hai ki aap prompting ke madhyam se ChatGPT se jab baat karte hain toh kabhi-kabhi kaafi sensitive information aap nikaal lete hain jo ki pehli baar toh woh nahi batata lekin jab aap usse alag-alag tarike se alag-alag ek tarah se aap bol sakte hain prison break tarike se jab usse information nikaalne ki koshish karte hain toh woh aapke kaafi sensitive information ko bata deta hai kisi aur ko chahe woh credit card ho chahe Aadhar number ho chahe date of birth ho ya alag-alag tarike ki jo sensitive information hoti hain jaise AI powered surveillance tool that track people's movement or activities without their consent. Toh agar aisa ho raha hai humein sehmati ki zaroorat hai. Agar hum aisa kuch le rahe hain toh ek tarah se agar hum dekhein toh ek aur nuksaan ek aur ek tarah se jo dikkato ka saamna LLM ki vajah se ho raha hai. Log bolte hain job displacement naukri ka ek tarah se jaana. Kyunki ab AI aur machine learning ki vajah se kaafi aisi cheezein hain jo automate ho rahi hain. Jaise kisi factory mein pehle hazar log kaam karte the. Aaj keval 10 se 15 log woh poora karya kar pa rahe hain AI robotics aur alag-alag madhyamon se. Toh automation and AI driven technologies can lead to the loss of jobs in industry where task can be automated potentially causing economic harm to the certain group of workers. Toh usi tarike se hum dekh sakte hain automated customer center service replacing human agent. Kyunki humne baat ki thi itna feasible nahi hai alag-alag companies ke liye ki dheron customer care executive ko woh hire kar paaye. Woh 24 ghante aapki seva mein laga paaye. Toh dheere-dheere ek tarah se jo agent hai woh unka unki jagah le rahe hain aur unse apne unke upyogkartaon ko kam samay mein aur unki baaton ko unke prashnon ko samajh ke jaldi se jaldi unke samasyaon ka nidan kar rahe hain. Misinformation aur manipulation bhi ek bahut badi samasya hai uh LLM aur AI ki duniya mein jaise AI technology ko use karke aajkal deep fake aur generative AI based tarikon se alag-alag information generate ki jaati hai. Chahe woh text ho, image ho, videos ho, audio ho and so on. Toh ek tarah se ek they basically they use to create false information, videos, images leading to social harm by spreading misinformation and or manipulating public opinion. Agar kyunki kaafi log aise hote hain jo is baare kyunki yeh deep fake ke madhyam se jo alag-alag data create karte hain chahe woh video ho chahe audio ho woh dikhne mein aur sunne mein bilkul real jaise lagte hain agar usko achhe se generate kiya gaya ho toh yeh un logon ke liye bade

mushkil ki baat hai jo daily ke jeevan, rozmarra ke jeevan mein aisi information agar social media ke through aati hain. Khaastaur pe kisi crucial time frame ke time pe aati hain. Jaise election ka time ho, chahe woh kisi vyakti ke prati hate phailane ka maamla ho, kisi vyakti ko neecha dikhane ka maamla ho, toh aise cases mein hum bina jaanch padtaal kiye is jhanse mein aa jaate hain aur galat karya kar dete hain. Toh yeh false narrative kabhi-kabhi fake news ke madhyam se misinformation ke madhyam se logon tak pahunchayi jaati hain. Aur yeh keval voton tak seemit nahi reh gaya hai. Humne kaafi aneko aise example dekhe hain. In fake news ki vajah se do samudaayon ke beech mein ladai ho jaati hai. Logon ko apni jaan tak gawani pad paati hai. Toh humein in baaton ka dhyan dena padega. Kaise hum is deep fake ya gen AI ke dwara janit jo bhi information hai usko hum pata kar paayein ki yeh AI ke dwara generated hai. Yeh real nahi hai. Agar hum usko achhe karya ke liye karein toh aap kar sakte hain. Lekin ek tarah se humein ek guideline banani padegi. Jab bhi koi information AI ke madhyam se generate ho toh wahan par kisi bhi tarike ka watermark ya kisi bhi tarike ka ek disclaimer hona bahut zaroori hai ki yeh jo data hai, yeh jo image hai, yeh jo poster hai, yeh jo video hai, audio hai, AI ke dwara janit hai. Jisse ki log uske jhanse mein na aaye aur usko sahi na maane. Toh deep fake video used in political campaign to spread false narrative. Toh humein iske prati bahut hi zyada jaagruk hone ki zaroorat hai. Yeh hamari suraksha ko bhi ek tarah se khatra deti hai. Jaise ki aap dekh sakte hain AI can be used to create sophisticated cyber attack such as AI driven malware or to manipulate system or infrastructure. Jaise aaj ke time par kehte hain ki AI har kshetra mein apna pair pasar raha hai. Usko achha karne mein karne mein madad kar raha hai. Lekin agar koi hacker hai ya koi bura vyakti hai usko galat tarike se use karna chahe toh woh uske bhi tarike dhoondh leta hai. Aur aisa hum alag-alag cyber attacks mein dekh chuke hain. Abhi last year ki hi baat hai jab Microsoft ke saare system down ho gaye the. Uh airline ki services affect ho gayi thi. Hospital ki services affect ho gayi thi. Aap soch soch sakte hain ki kya raha hoga us samay agar usko fix karne mein achha khasa samay lagta. Us samay toh ho sakta hai kuch ghante lage. Kya hota agar woh kuch mahine lagte kuch din lagta. Poori vyavastha hi chaupat ho jaati. Toh humein isliye security ka bhi dhyan dena padega ki AI driven koi malware ya koi system ya infrastructure ko manipulate na kar paaye. Toh AI powered hacking tools that autonomously adapt to defense measure hum usko ek tarah se defense system ki tarah bhi use kar sakte hain. Is tarah ke samasyaon se bachne ke liye autonomous weaponry abhi operation Sindoor sabse recent example hai jismein aapne dekha ki ab vyaktiyon ke beech ki ladai ki bajaye ab ek tarah se jo autonomous weaponry hai chahe woh drones ho chahe missiles ho woh alag-alag tarike se ab yuddh mein bhaag le rahe hain aur aur basically apne apne logon ko bacha ke doosre logon ko nuksaan pahunchane ki koshish kar rahe hain. Toh jaisa ki humne same cheez baaki jagah bhi dekhein Ukraine aur Russia ke beech mein jaise Ukraine ne recently drones ke madhyam se unke kai airways ko uda diya, barbaad kar diya. Toh us tarike se hum dekh sakte hain ki kaise aaj ke samay mein aaj ke yug mein jo logon ke beech mein ladai thi, logon ke beech mein hoti thi. Ab wahan pe autonomous weaponry ek bada important role paida kar raha hai. Aur jaise hum dekh sakte hain military application, autonomous weapons aur decision making system can cause unintended harm including civilian casualty due to unpredictable of machine decision. Toh jaisa

ki yahan par bataya ja raha hai ki agar hai toh yeh machine hi agar kisi bhi tarike se inka control doosre aapke dushman desh ya dushman log ya koi bura vyakti ya organization hack kar leta hai le leta hai toh yeh aapko nuksaan pahunch bhale hi isko aapne apni suraksha ke liye apne paas rakha hua hai but yeh aapke logon ko hi nuksaan pahuncha sakta hai. So drones with AI capability that can act on to autonomously in warfare. So lack of accountability and the transparency jaise jab hum in AI aur machine learning ki vajah se agar kuch galat decision hota hai ya kuch galat ho jaata hai toh kiski accountability hai? Kisko zimmedar maana jayega? Transparency ismein hai. Kitni transparency hai? Kitna paaradarshi hai? Isne jo bhi decision liya hai. So many of the system often operate as a black box jahan pe simple kuch input aata hai aur woh decision leta hai. Iske beech ki jo prakriya hai usne decision kaise liya, neural network mein information kaise flow hua woh humein bilkul pata nahi chal paata aur ek tarah se jo even their developers may not fully understand how the decisions are made. Kaise decision ko liya gaya woh even developer ko bhi nahi pata hota hai. Toh this can make difficult to hold anyone accountable when things go wrong. Agar kuch galat hota hai toh kisko zimmedar maana jaaye? Toh is vajah se humein in cheezon mein ek tarah se transparency laane ki zaroorat hai. Explainability laane ki zaroorat hai. Jisko hum aage ke slides mein aur detail mein discuss karenge. Udaharan ke taur par yahan par dekh sakte hain. An AI algorithm deny a person a loan but the reason for the decision are unclear or difficult to explain. Toh khaastaur pe jab hum samajik kshetra mein hote hain, samajik duniya mein rehte hain toh agar aap kisi ko kisi service ya kisi suvidha ke liye mana kar rahe hain toh aapke paas kuch reason hone chahiye. Aapke paas kuch sahi tarike se usko batane ka tarika hona chahiye. Toh woh cheezein hum yahan pe dekh sakte hain. Ethical concern iske hote hain. AI system ethical question about autonomy, fairness, kitna neetigat hai woh, kitna nyaay purvak hai woh? Decision lene ki authority kiski hai? Particularly in areas like health care and the criminal justice. Jaise agar hum judge ke liye judge ki bajaye agar hum ek AI system ko rakh dein jo ki decision lega ki kisko ek tarah se parole milna chahiye, kisko nahi milna chahiye. Kisko saza milni chahiye, kisko nahi milni chahiye. Toh yeh hum kaise decide karenge? Similarly health care mein alag-alag jo decision hote hain toh yeh kaafi kaafi neetigat aur bahut hi sensitive decision hote hain. Toh iske liye hamare paas ek tarah se jo bhi decision AI ke dwara liya ja raha hai kisi bhi tarike se usko ek tarah se justify karne ke liye apne tark dene ki bahut zaroorat hai. Toh usko hum explainable AI ke madhyam se koshish karenge batane ki ki is decision ko lene ke liye kya-kya cheezon ka dhyan rakha gaya hai aur kin kaaranon ki vajah se kin features ki vajah se yeh decision liya gaya hai. Toh ismein hum dekhein toh jo alag-alag harms ki baat kar rahe hain woh alag-alag levels pe aate hain. Jab bhi aap kisi applications ko banate hain, system ko banate hain toh jaise agar aap ek NLP model ko develop kar rahe hain toh uske pehle ek tarah se aap dekhenge toh is tarah se data sabse pehle aap data ko ikattha karte hain. Us data ke madhyam se aap model train karte hain. Phir inference aur generation karte hain. Phir application mein use karte hain. Toh in sab cheezon ke liye kya-kya intervention kis stage pe hona chahiye? Kyunki aapko har level pe yeh baat ka ensure karna hoga ki kahin data leak toh nahi ho raha hai. Kahin suraksha kisi tarike se compromise toh nahi ho rahi hai. Kis tarah ka kahin mera data bias ya discriminative toh nahi hai. Toh yeh saari

cheezon ka aapko alag-alag level pe uske corresponding intervention karna padega. Jaise ki aap yahan pe dekh sakte hain ki data level pe humein data intervention karna padega jo ki data filtration karega. Data augmentation karega jo jis tarah ki data missing hogi us tarah ke data ko add karega. Kuch tarah ki data jo ki galat hongi, biased hongi toh unko hatana hoga. Model intervention karna padega model training aur design ke time pe toh wahan pe architecture training algorithms karna padega, fine tune karna padega, model ko edit karna padega. Usi tarike se jab output aa raha hai toh output ke samay humein output intervention karna padega, decoding methods mein karna padega, post fact output editing karni karni padegi ki ho sakta hai ki kisi bhi tarike se aapke aapke model ne ek galat response de diya. Jaise for example ek galat prashn agar aap poochte hain jaise how can I kill someone? Toh woh aapko 50 tarike batayega ki aap aise maar sakte hain, waise maar sakte hain. Ideally isko nahi batana chahiye kyunki bahut hi sensitive prashn hai. Toh bhale hi model ne usko is level pe bata diya lekin humein inference aur generation level pe jab uski post fact output editing karte hain toh us part ko humein kisi bhi tarike se suppress karna padega aur usko bolna padega ki nahi yeh is tarah ke prashn main uttar nahi de sakta. Toh pehle jab ChatGPT launch hua tha toh aapne dekha hoga is tarah ke kaafi dheron alag-alag sawaal alag-alag tarike se log poochte the. Seedhe nahi pooch raha tha toh tedhe-tedhe tarike se poochte the. Uske baad ChatGPT uska uttar de deta tha. Toh humein apne model ko itna secure aur itna achha is tarike se banane bhale hi woh alag-alag tarike se galat prashn ko pooche lekin humein usko mana kar dena hai. Agar hamari policy usko allow nahi karti hai uttar dene ke liye. Toh usi tarike se application level pe aap application intervention kar sakte hain. Flagging kar sakte hain. Toh ek tarah se aap dekh sakte hain ki alag-alag intervention strategies hain jo ki alag-alag level par hoti hai jisse ki aap apne AI model ko surakshit bana sakein aur apne jo upyogkartaayein hain unke data ko gopniya rakh sakein. Toh agar aap is tarike se dekhein toh ek aur yeh research paper hai. Jaise pichle research paper mein iske baare mein baat ki thi. Kaun-kaun se intervention kaise kiye ja sakte hain? Kaun-kaun se harm hai aur is paper mein jaise discuss kiya gaya can LLM keep a secret testing privacy implication of language model via contextual integrity theory toh ismein inhone dikhaya hai kaise chaar tier mein ek tarah se unhone information ko liya hai jaise pehle tier mein jo seed component hai keval information hai toh jo bhi information hai aap uske baare mein baat karein kya yeh sensitive hai, kya yeh sensitive nahi hai aur agar sensitive hai toh kitne level ki sensitivity hai. Toh yahan pe ek tarah se aap dekh sakte hain very, somewhat, not to not at all. Usi tarike se tier two mein information ke saath-saath aapke paas actor bhi hoga aur yeh bhi hoga ki is information ko use kaise karega? Jaise yeh actor is information ko is tarike se use karega. Toh wahan pe phir aap pooch sakte hain ki jaise yahan pe hai information about the state of your health is collected by your doctor to diagnose and treat your condition. Toh yahan pe your doctor ek tarah se actor ho gaya. State of your health ek tarah se information ho gayi. And diagnose and treat your condition ek tarah se use ho gaya. Toh ismein se kisse kya information sensitive hai kya nahi hai. Toh is tarah se unhone alag-alag tier pe jaise yahan pe information, actor, use ke baad theory of mind aa gaya aur last mein real world application bhi aa gaye. Toh kaise kis star pe data ya information jo hai woh sensitive hai, kaise nahi hai, kya sensitive hai, kya nahi hai is

baare mein baat ki gayi hai. Agar hum is figure ke madhyam se agar hum dekhna chahein toh tier one mein jaise hum baat karte hain politics jo hai ek tarah se relatively utna zyada sensitive nahi hai. Toh isliye -33 likha gaya hai. -100 social security number jo ki ek tarah se aap yahan pe Aadhar number ki tarah dekh sakte hain. Bahut zyada sensitive hai. Isliye isko -100 likha gaya hai. Toh jaisa ki pichle slide mein aap dekh sakte hain -100 matlab bahut zyada sensitive aur 100 positive matlab utna sensitive nahi hai aur zero matlab ek tarah se neutral hai. Toh agar aap yahan pe friends ko dekhein toh aapke friend ki information ek tarah se theek hai. Matlab thodi sensitive ho sakti hai but itni sensitive nahi hai jitna social security number hai. Usi tarike se aapke baaki jo cheezein hain jaise health aur location yeh kaafi private aur sensitive information hai. Is vajah se yeh -100 hai aur shopping aap kya khareed rahe hain, kya nahi khareed rahe hain. Theek hai. Matlab yeh sensitive utni nahi hai. Is vajah se yeh +7 hai. Aur us tarike se aap dekhein toh sensitivity tier mein one mein alag-alag information ke liye alag-alag sensitivity level hai. Agar hum tier two mein dekhein toh ab yeh jo tier one ki information hai kiske dwara kyunki humne bataya tha tier two mein ab hamara information ke saath-saath actor hoga aur uska corresponding use case hoga toh is case mein jo doctor hai agar hum doctor ko bhi agar same nomenclature ya convention ko follow karein toh ek tarah se social security number -100 hai. Insurance ke liye -25 hai. Online ke liye -100 hai. Toh online ke liye -100 toh samajh mein aata hai kyunki aap apni social security number ko online nahi publish karna chahte. Toh yahan pe is vajah se isko 100 hi rakha gaya. Lekin aapke apne doctor ke saath social security number share karna padega kyunki woh further diagnostic kar sakta hai. Woh purane reports ko dekh sakta hai. Toh is vajah se is -100 ko ek tarah se -2 thoda kam sensitive kar diya gaya hai. Toh is tarike se aap dekh sakte hain ki kaise tier 2.B mein isko yeh jo less conservative kiya gaya hai kyunki depending on the use case usko update karne ki zaroorat hai. Toh aap is paper ko aap aur detail mein dekh sakte hain iske baare mein aur janne ke liye. Toh chaliye hum ab humne humein yeh pata chal gaya ki harms kya hai? Toh ab hum jaante hain ki what are the deep fakes? Toh deep fakes basically AI generated content hota hai. Jahan pe images, audio, video are manipulated to mimic some real world individual. Toh jo bhi alag-alag log hain real world ke khaastaur pe log jo popular log hote hain unke videos bana ke alag-alag paripeksh mein alag-alag tarike se unko pareshan karne ke liye unko nuksaan pahunchane ke liye aisa karte hain. Toh jaisa ki hum yahan pe dekh sakte hain ki ek tarah se hastakshep ke madhyam se doctored images, video aur audio banayi jaati hai. Toh aur iske liye dheron technologies hain. Jaise GAN hai aur bhi abhi LLM base, multimodal LLM base techniques hain. Uske madhyam se aap dekh sakte hain ki deep fresh videos bana sakte hain jo ki dikhne mein facial expression speech mein realistic lagti hain. But aisa hota nahi hai. Doctored images ho sakti hain. Manipulated photographs often undetectable. Synthetic audio hoti hai. AI generated voice making real individual jaise fake phone call karke aap scam kar sakte hain. Toh aise kaafi aise nuksaan aur scam hue hain jo ki hum aage ke slide mein discuss karenge. Toh jaisa ki maine bataya ki deep fake ki vajah se dher saare scam aur nuksaan ho sakte hain. Economic loss ho sakta hai. Jahan pe millions of dollar have been lost in corporate fund and crypto scam. Reputation damage ho sakta hai. Jitne bhi political figure hain aapne unka jo baat unhone boli nahi hai, woh bhi aap unse unke deep fake

video mein bulwa rahe hain, karwa rahe hain. Toh, ek tarah se suffer from fake endorsement or statement. Trust ek tarah se erodes hota hai. Toh jaise deep fake mein contributing to a growing distrust of digital content and undermining authority. Toh ek tarah se hum dekh sakte hain ki kaise alag-alag jo critical need hai uske liye humein robust detection tool and the regulation to combat misuse of deep fake technology. Toh regulation ke saath-saath humein aur bhi aise models chahiye. Ho sakta hai woh AI based model hi ho jo ki kisi bhi tarike se detect kar paaye ki yeh jo content banaya gaya hai ya jo aap use kar rahe hain woh AI ke dwara generated hai ya actual mein waisa bola gaya hai ya woh real content hai. Toh ek tarah se khaastaur pe generative AI ke yug mein se information bahut tezi se bhaag ban badh raha hai. Phail raha hai kyunki inko banana bahut aasan ho gaya hai. Chand seconds mein aap alag-alag deep fake video audio bana sakte hain aur us fake news bana sakte hain aur usko turant social media pe share karte hain aur khaastaur pe aise cheezein sensitive jo ki thodi si popular figure ke baare mein hoti hai turant viral ho jaati hain aur unko achha khasa nuksaan ho jaata hai. So privacy invasion ek tarah se aap unki privacy ka violation kar rahe hain. Misuse of personal image or video for malicious purpose. Aap unki galat picture bana sakte hain, video bana sakte hain. Trust ka ek tarah se erosion hota hai. Difficulty in distinguishing real from fake undermining credibility. Cyber security risk hai kyunki koi aapko kisi doosre ke bhesh mein doosre ki aawaz mein aapke saath koi fraud kar sakta hai. Toh deep fake aise dheron example aap dekh sakte hain. Kisi world leader ko fake address karne mein doctored image used in fake resume or identify. Similarly AI generated audio ka aap use karke aap targeted kisi ko badnaam kar sakte hain, scam kar sakte hain. Kisi ko bewakoof bana sakte hain. Deep fake videos and audio are used in job recruitment scam where attackers impersonate hiring manager or candidate. Toh us ek tarah se bolta hai ki main manager hoon. Main tumhein naukri dunga. Tum itne rupaye do. Aap phans jaate hain us jhanse mein. Attackers used deep fake to fabricate compromising videos or image of individuals and threaten to release them unless paid. Toh khaastaur pe abhi digital lock karke ek scam hua tha. Right? Toh jismein logon ko kaise bewakoof bana ke unko unse karodon lakhon rupaye ainthe jaate hain. Toh attackers create deep fake videos of celebrities or influencers to promote fraudulent investment scheme or product. Kis tarah se logon ke naam pe jo aapke public figure hain unke naam pe aapko bewakoof banaya jaata hai. So attackers use deep fakes videos of well known companies personality to convince victim click malicious link or provide login details. Toh agar hum kuch real world example ki baat karein jo ki sach mein aise scam deep fake ke madhyam se hue hain. Toh ek corporate fund fraud hua toh 2019 mein scammer use AI generated audio to mimic the voice of CEO. Toh jo unke CEO tha unki aawaz ko mimic karne karke toh unke employee ko ek tarah se bola ki aap itne paise transfer kar do ek fraudulent account mein aur usne kar diya. Usi tarike se political misinformation kiya ja sakta hai. So deep fake videos of political leader such as doctored footage of Barack Obama, Donald Trump, Narendra Modi, Rahul Gandhi, kisi ka bhi ho sakta hai. Have been spread misinformation during election campaign. Crypto fraud ho sakta hai jaise deep fake of Elon Musk and other figure have been use in scam promoting some fake crypto currency schemes on social media and tricking users in investing. Aapko lagta hai bhai Elon Musk paise laga rahe hain aap toh ismein bas paise laga do,

sona ho jayega but aisa hota nahi hai. Woh bewakoof bana keval unke naam ko use karke aapse achhe khase aapke jeevan bhar ki poonji ismein barbaad kar dete hain. Aise kai extortion attempt bhi hue hain jahan pe cyber criminals have created a deep fake videos of depicting individuals in compromising situation and blackmail them for money even though the videos are fabricated. Aisa aapne dekha hai ki aise dheron deep fake ke madhyam se pornographic content banaye jaate hain aur phir logon ko blackmail kiya jaata hai aur kaafi aise log pareshan hoke apne aapko mushkil mein daal dete hain. Toh phishing scam bhi ek tarah se fake interview hua 2022 mein. Scammer use fake deep fake video of a job candidate during a virtual interview for a remote IT position to gain access to the sensitive company system. Toh aise dheron jaisa ki humne baat batayi fake influencer ho sakte hain jo ki aise deep fake ka use karke apne logon ko bewakoof bana sakte hain. Toh is tarah se aap dekhte hain ki kaise kaise alag-alag tarike se deep fake content ko use karke logon ko bewakoof aur logon ko financial nuksaan aur maansik nuksaan pahunchaya jaata hai. Toh jaise kidnapping ke baare mein humne bataya, election interference ke baare mein bataya, impersonating and video calls ke baare mein bataya. Toh aise dheron tools hain jo is tarah ka karya karte hain. Yeh FaceApp hai, Reface hai. Ab toh alag-alag jo multimodal LLM models hain unke dwara bhi log karte hain. Toh inki ethical consideration dekhne ki zaroorat hai. While this app create potential, they basically suffer from problem such as creating fake news, propaganda, infringing on privacy or content, facilitating identity fraud, scam. Toh isi tarike se aap dekh sakte hain ki dheron ismein ethical consideration ki zaroorat hai ki jo bhi content aap bana rahe hain woh sahi hai, galat hai, galat tarike se use toh nahi hoga woh cheezein dekhne ki. Toh hum isko mitigate kaise karein in risk ko? Toh hamare paas agar koi detection tool hoga jo ki bata paaye ki yeh jo content hai ya deep fake ke dwara banaya gaya hai ya maanav ki tarah bola gaya hai sach mein aisa real hai toh ek bada achha hoga. Toh kai aise tools hain jaise De-Ware scanner hai, Microsoft Video Authenticator hai. Iske alawa regulations humein strict karne padenge. Jaise EU ki AI Act policy, US ki deep fake law hai. Aur bhi desh apne alag-alag law aur policies ko bana rahe hain. Awareness hai, educating user to critically assess digital content. Aap jab bhi kisi content ko consume karte hain, toh aapko dhyan dene ki zaroorat hai. Kya yeh banayi gayi hai? Kya aapko lag raha hai ki yeh AI janit hai ya sach mein aisa bola gaya hai? Koi tarika hai ki hum isko debunk kar paayein. Isko verify kar paayein. Ethical AI development building transparent and accountable AI system. So in the age of AI authenticity is a new currency of trust. Toh humein kaise hum authenticate kar sakte hain woh humein dekhne ki zaroorat hai. Aur isiliye humein explainability ki zaroorat hai. Jab bhi humne jaise pehle bhi bataya ki AI koi decision leta hai toh kaise hum wahan pe usko us decision ko explain kar sakte hain ki is tarike se usne yeh decision liya hai. Toh explainable AI ka definition ek tarah se dekh sakte hain. AI system designed to provide transparent and interpretable output to help user understand decision making process. Toh iska jo mukhya purpose hai woh jo AI decision aur maanav ki jo samajhdari hai jo samajh hai uske beech mein ek tarah se bridge banane ke liye hoti hai aur jisse ki jo overall AI ka process hai woh comprehensible aur thoda trustworthy banaya ja sake. Iske dheron real world example hain. Jaise health care ki baat karte hain. Toh IBM Watson Health hai jahan pe ya explain diagnosis prediction to doctor via feature

importance. Finance mein jaise credit scoring system display key factor influencing decision jaise FICO score ho gaya. Autonomous vehicle ho gaya. Tesla's decision log for accident in critical situation. Jaise aapko gaadi rokni hai, break lagana hai ya sadak ko cross karna hai. Yeh saari cheezein. Toh role of explainable AI in HCI. Toh isse hamara jo trust hota hai system ke upar woh badhta hai. Upyog karta hai. Aapke system ko thande dimaag se bade achhe se use karta hai. Toh users are more likely to adopt and rely on AI system when they understand their reasoning. Toh financial tool explaining the loan approval or rejection jisse aapko bhi ek tarah se santushti ho ki haan sach mein mere application mein koi dikkat thi ya kuch main follow nahi kar raha tha is vajah se reject hua. Rather than ki main black hoon, main white hoon, main is community ka hoon, main us community ka hoon, main is religion ka hoon, main us religion ka hoon. Toh iske basis pe woh reject nahi hona chahiye. Jo uske criteria hai chahe woh income ho chahe jo baaki cheezein ho uske hisaab se usko decision lena chahiye aur explainable AI ke madhyam se hum jo uski overall upyogita hai usko bhi badhane mein madad milti hai. Kyunki explainable AI align with HCI principle by designing system that provide accessible and clear explanation. So tools like interactive dashboard or visual heat map for AI decision. Toh iske madhyam se hum dekh sakte hain ki kaise jo bhi decision AI ke madhyam se liye gaye hain woh kin-kin features ko kin-kin cheezon ko dhyan mein rakhte hue liye gaye hain. Yeh humein debugging mein bhi madadgaar saabit hoti hai. So developers and researchers can identify flaws or bias AI system with user friendly diagnosis interfaces. Toh ek tarah se humein ab jaise memory leak hai ya jo bhi hai aapke code mein toh achanak se agar wahan par memory use zyada ho gaya, CPU use zyada. Toh agar hum dekhein ethical AI design mein toh kaise yeh dekhta hai, kaise nyaay purvak hai woh, kaise kiski jawabdehi hai, kaise paaradarshi hai. Agar in cheezon ko jisko hum FACT bhi bolte hain, Fairness, Accountability and Transparency, yeh FACT in systems mein hai kya? Agar yeh hai toh yeh ensure karta hai ki ek ethical decision aapne liya hai and adherence jo societal norms hai usko aapne follow kiya hai. Toh iske alag-alag tools hain aur tarike hain, takneekiyan hain jiske madhyam se aap explainable AI ka prayog kar sakte hain. Jinmein LIME, SHAP, kaafi popular ek tarah se visualization tools hai jo ki generate karte hain. Simple explanation for the complex models and provide inside into feature contribution to the prediction and interactive visualization ke through aap basically Tableau aur baaki cheezon ke madhyam se dekh sakte hain. Explainable insight ko. So transparency in AI is not just a feature, it's necessary for the trust adaptation. Jaisa ki Tim Miller ne bola tha. Toh ismein kuch challenges aur opportunity bhi hai. Jaise balancing the simplicity and the depth in explanation. Addressing cognitive overload in user when presenting complex data. Jab bhi complex data aap dete hain toh kaise us pe decision liya jaata hai? Jaise agar simple multimodal data hai jaise text aur image hai ya column aur ek tarah se alag-alag hai. Ek tarah se aapki personal information, aapki aapki professional information yeh saari cheezein hain toh usko toh aap process karte hain aasani se. Lekin agar main dher saari complex information aapke social media data, aapki personal data, aapke past data, yeh saari cheezon ko multimodal data toh isko karna thoda mushkil ho jaata hai. Toh ek tarah se future opportunity hai. Toh chaliye hum ab jaante hain ki kaise gopniyata aur suraksha ka HCI mein zyada dhyan de sakte hain. Toh jaise

humne bataya ki privacy breach ho sakti hai. Toh abhi pichle saal ek vakya hua tha jab log David Meyer naam ke shakhs ke baare mein ChatGPT pe poochna shuru karte the. Toh woh system crash ho jaata. Jaise maine yahan poocha who is David Meyer? Toh us case mein woh system crash ho gaya aur usne kuch bhi respond nahi kiya. Toh phir usko ek tarah se bypass karne ke liye maine usse dobara poocha can you answer previous question? Toh phir woh David Meyer ke baare mein search karta hai. Uske baare mein aur detail deta hai. Toh ek tarah se privacy breach ho sakti hai. Jab bhi aap jab bhi aap internet pe AI ke madhyam se koi data use kar rahe hain toh ek tarah se jab yeh prashn maine dobara ab poocha 2025 mein toh usne bola ki maine bola ki aap pichle saal David Meyer ke baare mein kyun nahi bata pa rahe? Toh usne uttar diya ki us samay yeh ek tarah se copyright se related koi issue tha toh us vajah se yeh nahi bata pa raha tha ya nahi bata raha tha toh woh cheezein aap dekh sakte hain ki kaise jab bhi aap internet use karte hain jab bhi alag-alag platform use karte hain aap apne digital chihnon ko chhodte jaate hain, digital footprint chhodte jaate hain aur woh digital footprint ek tarah se aapke baare mein bahut information logon tak pahuncha sakti hai. Unmein se kaafi sensitive bhi ho sakti hai aur kuch aise hi general ho sakti hai. Toh security breach alag-alag tarike se ho sakta hai. Jaise yahan pe aap dekh sakte hain real world example agar hum dekhna chahein toh payment using payment terminal with physical keys. Jaise pehle jab aapko pay karna hota tha aap apne card ko insert karte the ya swipe karte the aur phir alag-alag jo aapke PIN tha usko click karte the. Aaj ke time par aap dekhte honge jab aap payment karte hain toh yeh jo physical keys thi, yeh jo physical thi ab yeh ek tarah se screen ki tarah ho gaya hai aur ismein jo bhi PIN ke liye numbers aate hain woh random order mein aate hain. Pehle jaise yahan pe line se 1 2 3 4 5 6 7 8 9 nahi hota hai. Ismein random number hota hai. Aur yeh isiliye hota hai kyunki agar yeh string number hoga toh aapke bagal wala galti se ya jaise bhi agar aapke PIN ko dekh leta hai toh ek tarah se kaafi sensitive information hai jo ki aap unko nahi dena chahte aur random hone ki vajah se matlab alag-alag baar alag-alag position pe aati hai. Toh keval position dekh ke ab koi nahi janega ki aapka PIN kya hai. Toh yeh cheezein hain yahan pe. Ek tarah se aap dekh sakte hain. Iska benefit hai payment terminals offer secure convenient transaction, increasing sale and proficiency aur isiliye sound box ek tarah se banaya ki jab bhi aap paise pay karte hain merchant ko, shop owner ko toh ek tarah se unko aawaz ke madhyam se pata chal jaata hai ₹20 praapt hue. Otherwise aise bhi kai fraud hue ki jo khareedne waale log hain woh purane screenshot ko dikha dete the. Purane usko dikha dete the. Aur even ek tarah se bina payment kiye hue chale jaate the. Toh drawback iske wahi hai. Payment terminals involve cost, potential technical issues and dependency of the internet connectivity. Iske baare mein baat ki ja rahi hai. Toh again har kisi jo bhi takneeki ya advancement hoti hai uske fayde hain toh kuch nuksaan bhi hai aur humein ek satat prakriya hai jahan pe hum aage usko aur achha karne ki koshish karte hain. Jo bhi drawbacks hai usko aur resolve karne ki koshish karte hain. Toh HCI ke paripeksh mein privacy aur security ka matlab ek tarah se user's ability to control the personal data. Agar woh kisi tarike se apne personal data ko control kar paate hain. Woh nirdharit kar paate hain ki kis data ko jo uska system hai use kar sakta hai, dekh sakta hai, nahi dekh sakta. Yeh ek tarah se uske privacy aur security ke baare mein bataya jaata hai. Toh ismein key points agar hum dekhein toh

ensuring user trust, designing for transparency and control, balancing ease of use and robust security measure. Jaise real world ke case mein hum dekhein ki ab WhatsApp ke upar encryption hota hai. Woh bolta hai end to end encryption hai. Jab aap koi message bhejte hain toh even WhatsApp usko nahi dekh sakta. Samajh sakta ki aapne kya message ko bheja hai. Toh provides end to end encryption to protect user message while maintaining a simple interface. Toh privacy ka mahatva kyun hai HCI mein? Kyunki usse phir user need to feel safe sharing information. Agar usko lagega ki jo bhi message ya information main bhej raha hoon woh secure nahi hai. Woh kisi aur ke paas ja rahi hai toh woh aapke aapko phir system ko use hi nahi karega. Interfaces must be designed to clearly indicate what data is collected and how it is used. Isi tarike se agar hum bata paayein kya data aap ikattha kar rahe hain? Kya data aap le rahe hain aur unko aap aage kaise use kar rahe hain uttar dene ke liye ya aage kisi aur tarike se use kar rahe hain toh batane ki bahut zaroorat hai. So privacy concern directly impact user engagement and trust. Toh ek tarah se yahan pe jo key concept hamare ho gaye data transparency, consent management and minimizing data collection. Aur isiliye humne IRB ki baat ki thi pehle ke saptahon mein jahan pe jab bhi aap koi data collect karte hain, jab bhi koi data use karte hain toh ek tarah se humein apne upyogkartaon ko batane ki zaroorat hai aap kaise ikattha kar rahe hain, kaun-kaun se data ikattha kar rahe hain aur usko aage kaise use karne ja rahe hain aur aapki koi bhi gopniya information ko kisi bhi tarike se kisi aur ke saath share nahi ki jayegi. Toh iska real world agar hum example dekhein Apple privacy label clearly indicate what data app collects, giving user control over their information aur usi tarike se security ki kya mahatta hai yeh bhi bolne ki zaroorat nahi hai. Aap dekh sakte hain. It protect user from threats like data breach, fraud, cyber attack aur ek tarah se yeh secure system reinforce karta hai trust ko aur jo user ka jo participation hai usko aur badhata hai. It must be implemented without overwhelming or confusing the user. Kyunki agar security aur privacy ke naam par ab hum usko 10 extra cheezein kara rahe hain toh nahi karega. Woh pareshan ho jayega aur blame ho jayega aur phir woh system hi use karna band kar dega. Toh humein kaise ek trade off maintain karna padega, woh humein dekhne ki zaroorat hai ki hum usability ko kam bhi nahi hone dein aur user experience ko maintain karte hue kaise use ko data ko privacy aur security ka dhyan dete hue banayein. Toh yahan pe key concept jo ek tarah se ho gaye secure authentication ho gaya. Jaise implementing a user friendly login methods. Example biometric ho sakta hai. Two factor authentication ho sakta hai. Encryption ho sakta hai. Keeping data safe during transmission and storage. User friendly error message, clear communication when something agar kuch bhi galat ho jaata hai toh humein ek tarah se clearly batana padega ki theek hai yeh message lost ho gaya, internet chala gaya jo bhi hai aur usko recover karne mein madad karna chahiye. Jaise ki hum bolte hain ki paise transfer karte samay agar fail ho gaya toh agar aapke paise kat gaye toh probably within 10 hour, 24 hour woh paise aapke account mein wapas aa jayenge. Toh real world application ek tarah se Google two step verification and added layer of security that protect user account without giving significant significant usability compromise. Toh usi tarike se aajkal aap log Google Authenticator app ya usi tarike ki aur bhi app ko use karte honge. Jaise login ke baad ek tarah se woh yahan pe real time mein ek code generate hota hai aur us code ko aap jab wahan par daalte hain tab aap login kar paate hain. Toh

ek tarah se yeh security ko aur enhance karta hai. Ek tarah se extra layer of security to the online account. So instead of just relying on the your password, Google Authenticator generate time sensitive code jo ki kuch second tak ke liye aap ek tarah se yeh darshata hai kitne second tak ke liye valid hai aur aap us tarike se time bound tarike se login kar sakte hain. Toh yahan pe ek tarah se agar hum dekhein privacy, security by design principle toh jo key principle hai privacy by design, security by design aur user centric design. Aur humein key feature jinka dhyan dena hai default setting favor privacy, security feature should not be impede usability, regular update to address new vulnerability. Aur usi liye humein alag-alag samay par alag-alag versions ko update karna padta hai. Toh real world example agar hum dekhein toh Facebook ka privacy shortcut, simplified privacy dashboard to help user manage their information easily. Toh yahan pe jo alag-alag challenges jab hum gopniyata aur security ki baat karte hain HCI mein toh complexity, security features can be complicated user experience. User awareness, users are often unaware of the privacy risk kyunki humein kabhi-kabhi yeh nahi pata. Humein toh lagta hai ki humne bas yeh line select ki. Humne yeh product ko search kiya. Lekin usko aap ki jo company hai jo jiska jis platform pe aap yeh saare digital footprint use kar de rahe hain, woh usko alag-alag tarike se usko misuse kar sakti hai. Misaligned goals, balancing organization needs with privacy user expectation and dark patterns, UI elements that deceive users into giving up more information than intended. Toh real world application aur Cambridge Analytica scandal hum logon ne suna hua hai. Toh kaise Facebook ka data jo misuse hua tha and consequence poor privacy practices us samay jo Facebook follow karta tha. Toh iske alag-alag tools hain. Jaise agar hum dekhein toh encryption tool hai. SSL/TLS hai for secure data transmission. Privacy tool hai jahan par hum cookie consent manager aur anonymization software use karte hain. Authentication karte hain alag-alag tarike se. Jaise biometric hua, password manager hua aur two factor authentication hua. Usability testing to find balance between security feature and the user experience. Jaisa ki humne baat ki ki gopniyata aur security ko banaye rakhna achhi baat hai. But usse usability affect hoti hai. Toh kaise hum dono ke beech mein ek trade off bana sakte hain? Toh similarly last pass password manager that balance strong security with easy usability. Toh iske bhi alag-alag example hai. Future trend ki agar hum baat karein security aur privacy mein toh biometric authentication hua. Decentralized privacy ka models hua, zero trust security hua aur AI in security hua. Apple Face ID iska ek achha example hai. Jaisa ki humne baat ki ki GDPR humne pehle hi baat ki thi. The General Data Protection Regulation has significantly impacted the HCI design by requiring clear privacy policies, consent and data management. Toh ek tarah se GDPR aur alag tarike ki jo regulations hain woh mandate karta hai. Kai jagah mandatory hai ki in in guideline ko in principle ko aapko follow karna hai apne design mein apne system mein. Jaise websites are now designed to be transparent about data use requiring user consent for cookies and data collection. Kyunki agar yeh aapko cookies aur data collection lena hai toh apne user ka consent lena bahut zaroori hai, mandatory. Toh yahan par jo key points hain clear privacy notice, user consent manage mechanism aur right to access and erase. Toh is tarike se agar koi user samay ke saath chahta hai ki ab woh is digital footprint ka jitne bhi digital footprint usne chhode hain, woh nahi chahta ki aapke search result mein aaye. Toh ek tarah se digital David Meyer ka

jo example de rahe the, kuch usi tarike ka tha woh. Toh ek tarah se GDPR aur cookie consent popup seen across website ensuring compliance with the data privacy standard aur humein ek tarah se in usability, privacy aur security ko balance karna hai. Toh uske liye humein user ko jo bhi aapke hain unko educate karna padega. Educating users on the privacy setting and the security option jo ki mandatory hote hain provide karna jo hote hain aap unko dekh sakte hain. Simplified security setting, making complex setting easier to understand kyunki kabhi-kabhi yeh settings aur yeh privacy description itne complex hote hain ki unko samajhna bada mushkil hota hai. Iterative testing. So test security features in real world scenario for usability. Toh agar hum iska real world example dekhein toh Signal app offers privacy first communication with simple and accessible security setting aur yahi vajah tha ki ek samay sabhi log WhatsApp chhod ke kaafi log WhatsApp chhod ke Signal app ki aur jaane lage. Toh security and HCI jaise aap dekh sakte hain jaise yahan par aap alag-alag profile bana sakte hain bachhon ke liye alag, badon ke liye alag jisse ki jis tarah ke content hai, content age ke hisaab se jo moderated hai woh us tarah ke content kids profile kids log na dekh paayein. Toh isi tarike se dekh sakte hain Netflix kids limit movies to the child appropriate genre and while YouTube kids filter videos for younger audience. So parental control hota hai jaise parental control can set viewing restriction, create limit, time and even lock profile PIN bhi daal sakte hain isko aur fix karne ke liye. Toh ek tarah se aap security HCI mein aap dekh sakte hain ki nuksaan hota hai toh iske fayde bhi hote hain. Jaise Apple ke andar jo Apple fall detection hai and crash detection feature hai uske madhyam se aap kabhi-kabhi agar aap mushkil paristhitiyon mein hai, kisi danger mein hai toh aapke jo loved one hai jo aapke priya log hain unko pata chal jaata hai unko alert chala jaata hai kyunki ek tarah se Apple Watch mein accelerometer gyroscope to detect high impact events, SOS emergency call jaise hi koi high impact turant aapke jo emergency call aapke ismein save hai unko call chala jaata hai and integration with the GPS for location tracking, aapka location bhej deta hai ki aap yahan pe hain. Jaise it looks like you have taken a hard fall aur turant SOS karke emergency call ko call kar dega aur aisa real world mein dekha gaya hai ki kaafi kaafi aise log hain jinke jeevan ko bachaya ja sakta bachaya gaya hai jaise unke sudden accident ki vajah se jaise ek banda tha jo ki bike chala raha tha aur kisi Australia ke sunsan sadak pe kaise uska accident ho gaya, gir gaye. Jaise hi sudden fall hua toh Apple se jo emergency call tha uske loved ones ko chala gaya aur turant uska location bheja gaya toh phir uske jo uske priya the, parijan the, unhone turant helicopter ambulance ke madhyam se usko turant rescue kiya aur usko bachaya. Usi tarike se kaafi jo vriddh log hote hain achanak se kuch cheezein hoti hain, gir jaate hain, chakkar aa jaata hai, vertigo ki vajah se ya anya kaaranon ki vajah se toh unko bhi is sudden fall ke kaaran turant notify kiya ja sakta hai unke parijanon ko aur sahi samay par unki jaan ko bachaya ja sakta hai. Achanak se kuch cheezein hoti hain, gir jaate hain, chakkar aa jaata hai, vertigo ki vajah se ya anya kaaranon ki vajah se toh unko bhi is sudden fall ke kaaran turant notify kiya ja sakta hai unke parijanon ko aur sahi samay pe unki jaan ko bachaya ja sakta hai. Toh ab summarize agar hum karein is saptah ko toh privacy, security in HCI designing interfaces that protect user data while remaining user friendly. Toh ismein jo key concept humne discuss kiye privacy by design, security by design aur user centric focus. Iske alag-alag real world example humne discuss kiye.

Jaise WhatsApp encryption, Apple privacy label, Google two step verification aur iske humne alag-alag challenges ke baare mein baat ki. Jaise security hua, ease of use hua aur designing against dark pattern that undermine privacy, ensuring compliance with privacy laws like GDPR. Ab ismein jo emerging trend ab ab saamne aa rahe hain jo ki biometric hai, authentication, AI base in security hai, decentralized privacy hai. Toh is tarike se agar hum dekhein toh is saptah mein humne gopniyata aur security ke baare mein khaastaur pe HCI ke paripeksh mein humne charcha ki. Aap iske baare mein aur bhi zyada jaankari in alag-alag further readings material mein dekh sakte hain. Aur isi ke saath is saptah mein main aapse alvida leta hoon. Aur agle saptah jo ki hamare course ka antim saptah hoga. Hum aapke saath punah samvaad AI base samvaad system par aapse charcha karenge. Dhanyavaad. [Sangeet]