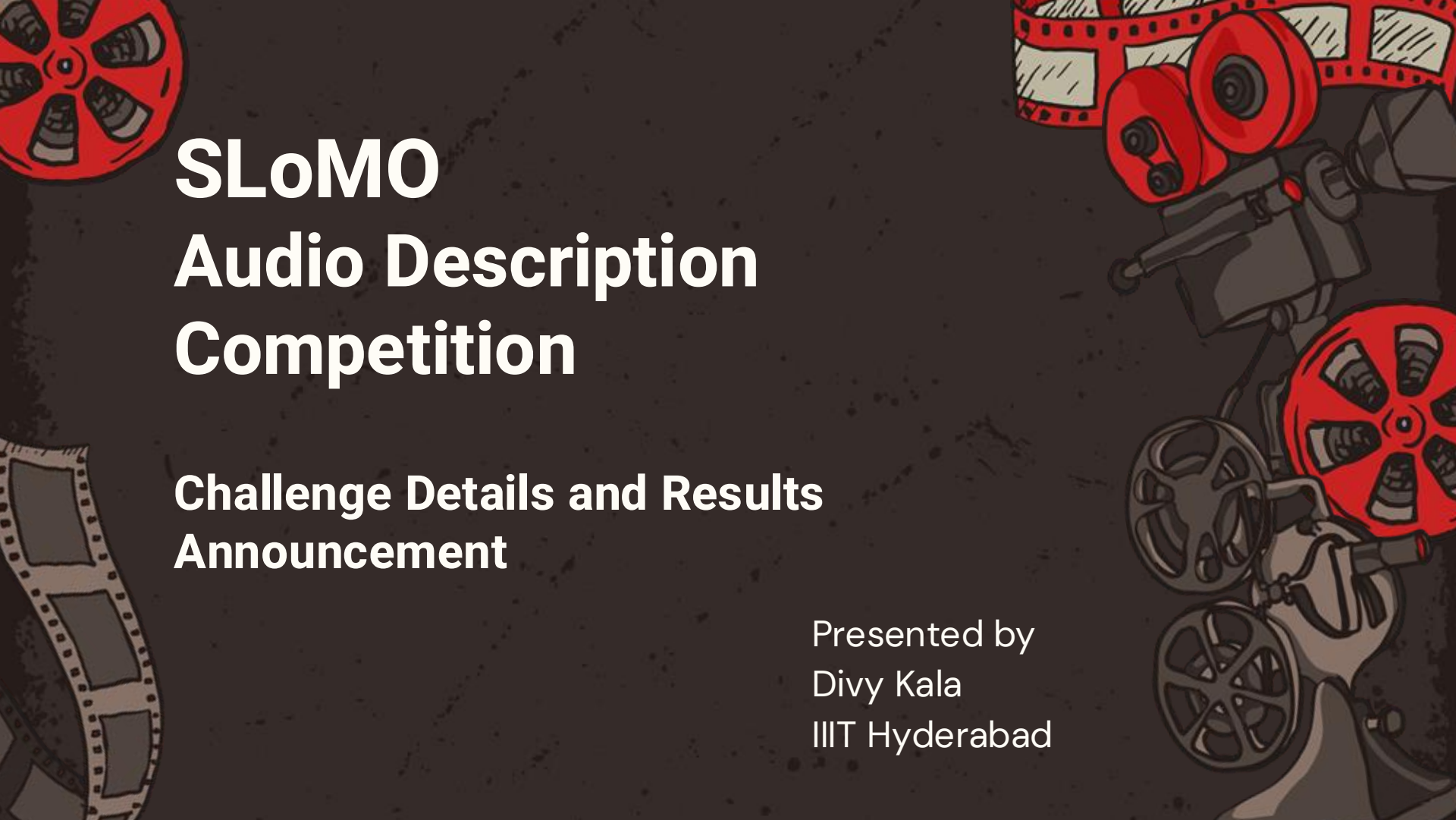


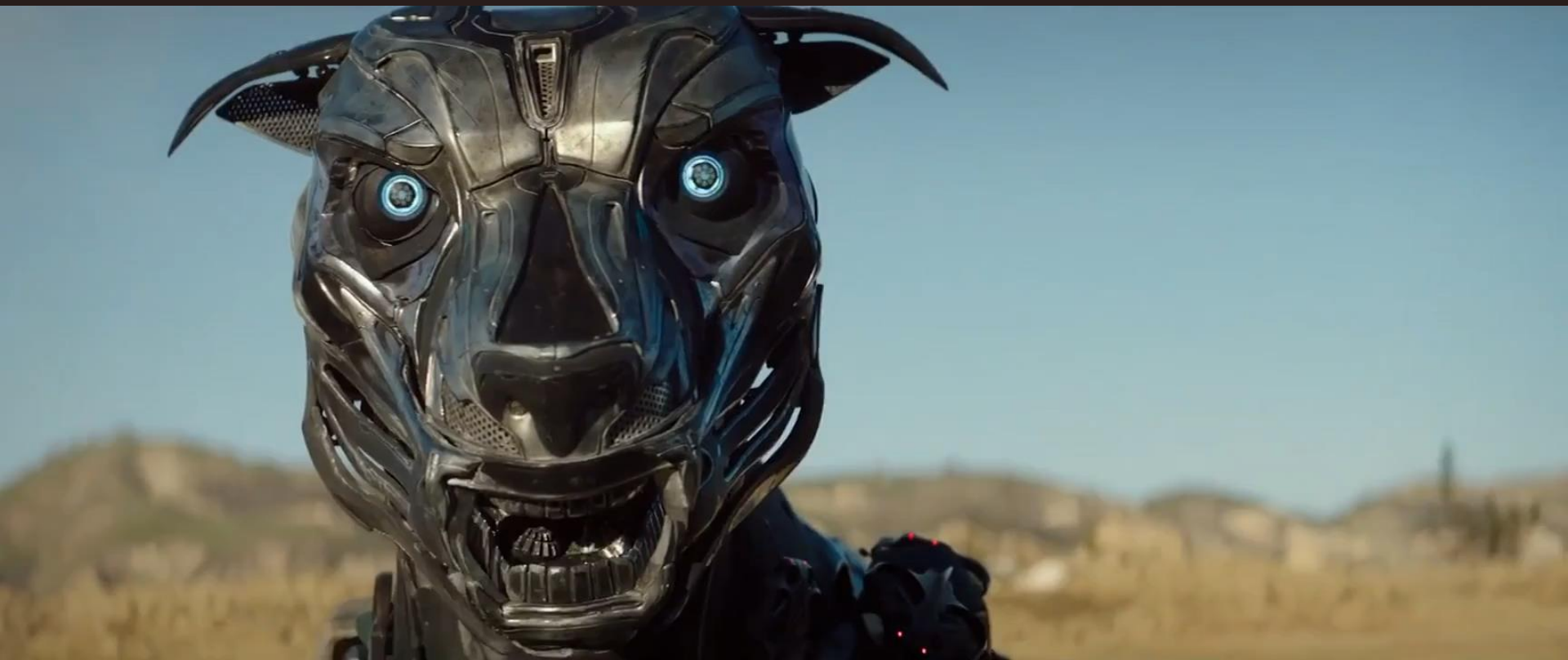


SLoMO Audio Description Competition

Challenge Details and Results Announcement

Presented by
Divy Kala
IIIT Hyderabad

What is Audio Description (AD)?



AD is really an art...

- ...and therefore subjective. Even professional audio describers disagree

- Barbie UK (11:10 – 11:30)

- <Silence for music>
- At Barbie's fairy-lit house
- Barbies wear glittery outfits
- They do a synchronized dance
- A barbie in a wheelchair spins



UK



US

- Barbie US (11:10 – 11:30)

- *At night*
- *In the cul-de-sac*
- *Outside* Barbie's dreamhouse
- Dances with other Barbies
- *Disco lights flash and swirl over them*
- <Silence for music>
- The *curvy lawyer barbie* carves the air with her arm
- A barbie in a wheelchair spins in a circle

What makes a good AD?

- Do's
 - Should be synchronized with the visuals
 - Should be objective, no interpretations (*beautifully decorated*)
 - Should help users *appreciate* the visuals of the media, and *understand* it (the two primary goals of audio description)
- Don'ts
 - Should not speak over dialogs or important audio
 - Should not state what is already understood from dialogs and sound
 - Should not over-describe (*let the movie breathe*)
 - Should not spoil content before it is revealed

How do we evaluate AD?

- CIDEr
 - Problem: Because ADs are subjective, a single reference can't be the only ground truth
- ADQA
 - Designed to check whether the generated AD actually helps in *narrative understanding* and *visual appreciation*.
- But ADQA alone ignores style and brevity → **CIDEr** keeps outputs close to how professional describers write

ADQA Questions

Narrative Understanding (NU)

How did the siblings manage to overcome the witch's scheme?

Who was responsible for spreading rumors about Hansel and Gretel's mother in the village?

What significant personal connection to the cabin do Hansel and Gretel discover?

Why did Hansel and Gretel earn a reward?

Visual Appreciation (VA)

What physical gesture does the man make towards the little girl?

How are the men whom the sheriff is facing characterized?

What specific object is Hansel's foot resting upon?

Towards which specific area within the bar does the man proceed?

Challenge Setup

- Task
 - **Input:** film clip + soundtrack + subtitles + character reference images + annotated time intervals
 - **Output:** one AD sentence per interval, within that interval's word budget
- Tracks
 - Main track (big models >10B params)
 - Special track (smaller models <10B params)
- Datasets
 - CMD-AD
 - Train: 1332 movies / Public Test: 98 / Private Test: 44 new movies
 - SF20K
 - Full-length short films. No official ADs exist for these films
 - Public Test: 17 films
 - Private Test: 26 films written by professional audio describers

Challenge Metric

- **ADQA**
 - gemini-3.1-flash-lite answers questions about the clip using only the generated ADs and dialog
- **CIDEr**
 - n-gram agreement with the professional reference ADs
- **AD Duration Check**
 - If an AD is too long for its gap, we delete the AD (qualifying check, but quite lenient)

$$\text{AD score} = 0.6 \times \text{ADQA} + 0.4 \times \text{CIDEr}$$

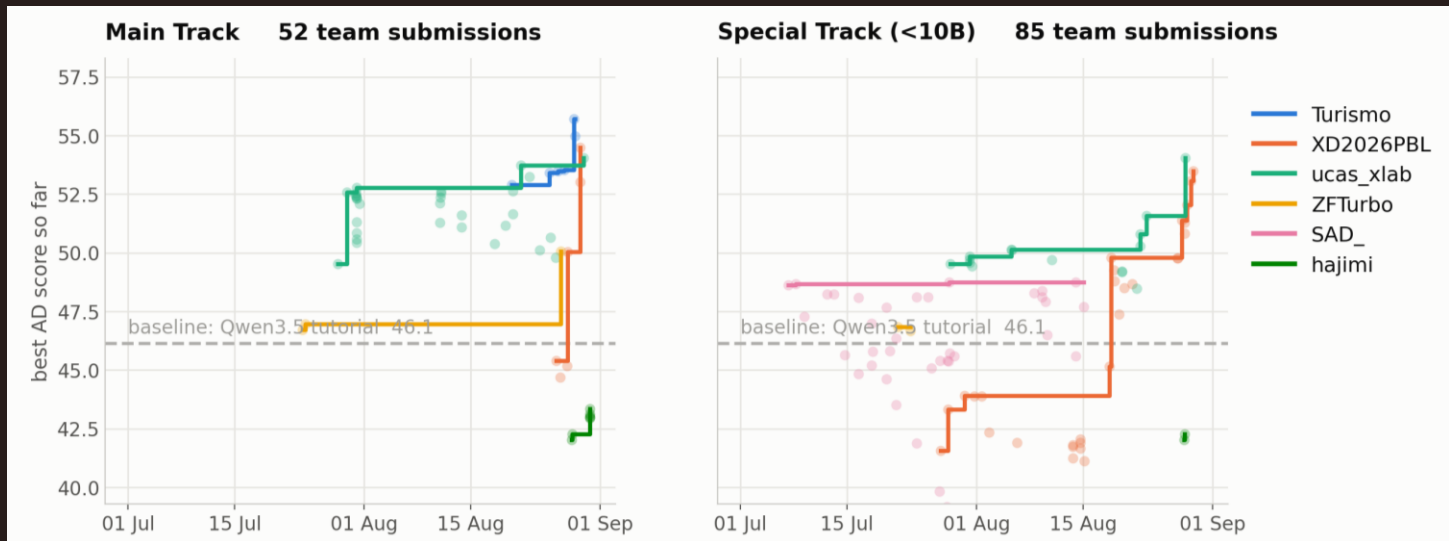
$$\text{final score} = \text{mean}(\text{AD score on CMD-AD}, \text{AD score on SF20K})$$

Public Phase of the Competition

- 1 July – 30 Aug
- Teams: 6
- Total submissions: 167

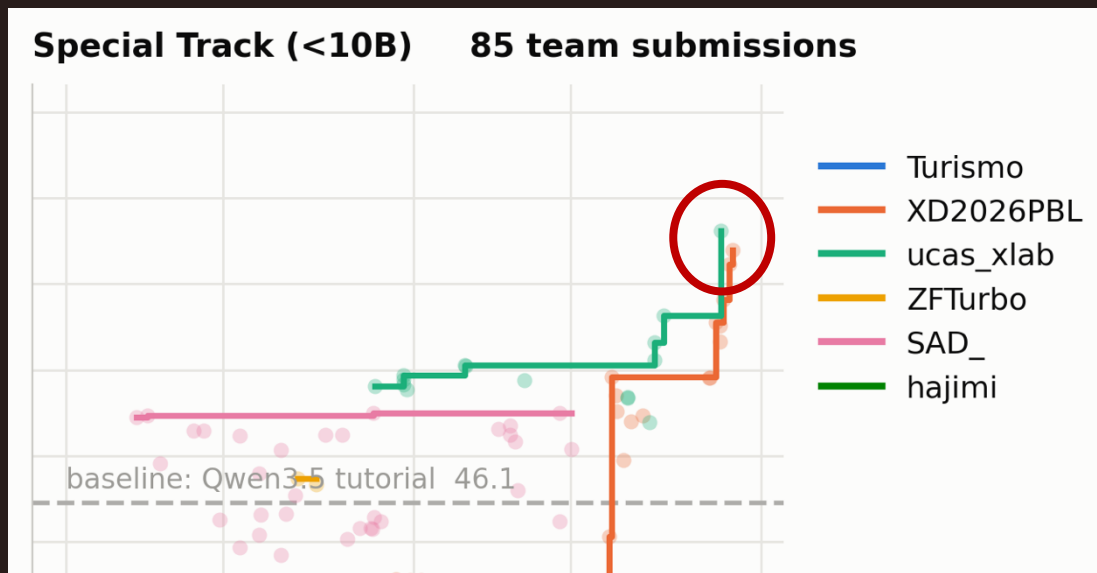
Public Phase of the Competition – Special Track

- Steps=personal best, dots=individual submissions



Public Phase of the Competition – Special Track

- Steps=personal best, dots=individual submissions



Private Phase of the competition

- 14 Aug – 30 Aug
- Teams: 6
- Total submissions: 30 (scores were not revealed)

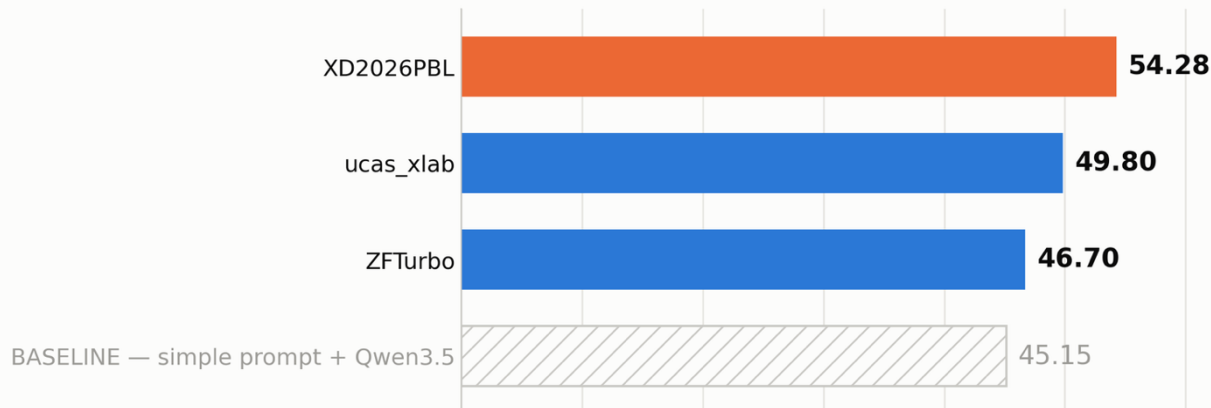
Special Track - Winners! Congratulations!

Special track	Teams	Prizes	
		Cloud GPU credits	
	XD2026PBL	\$500	
	Ucas_xlab	Sponsored by:	
	ZFTurbo	 MOHAMED BIN ZAYED UNIVERSITY OF ARTIFICIAL INTELLIGENCE	

Final Private Phase results - Special Track

Special Track (<10B params) — Private Test Leaderboard

higher is better · private test, best submission per team · hatched grey = organizer baseline



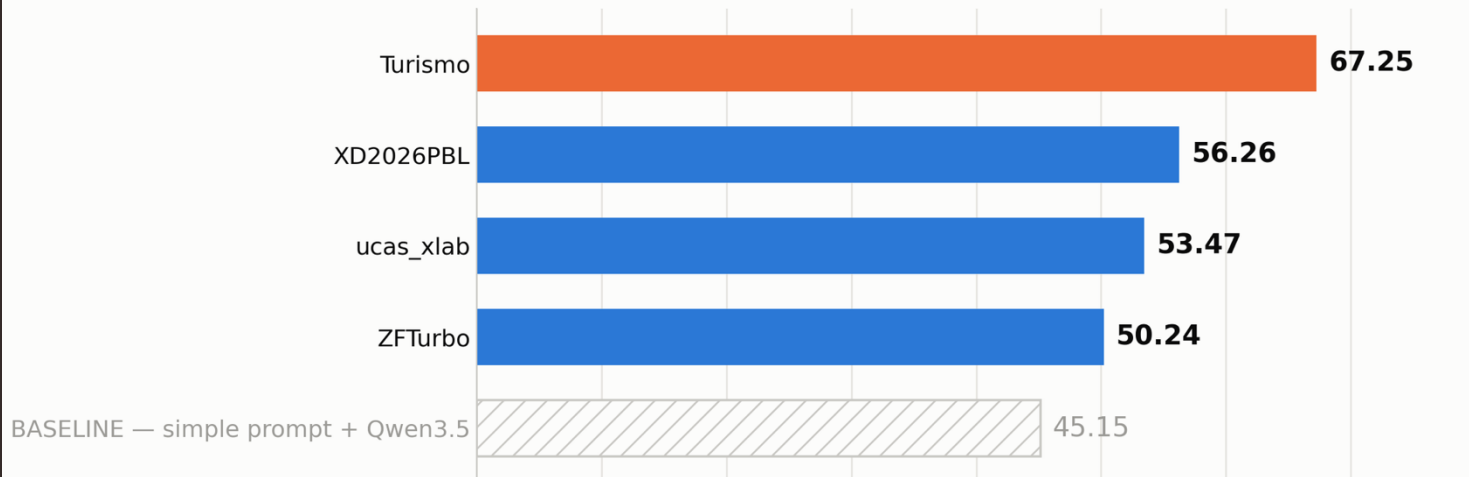
Main Track - Winners! Congratulations!

Main track	Teams	Prizes
	Turismo	Cloud GPU credits \$500
	XD2026PBL	Sponsored by:  MOHAMED BIN ZAYED UNIVERSITY OF ARTIFICIAL INTELLIGENCE
	Ucas_xlab	

Final Private Phase results – Main Track

Main Track — Private Test Leaderboard

higher is better · private test, best submission per team · hatched grey = organizer baseline



ADQA vs CIDEr

- CIDEr and ADQA correlated well with each other
- Except one submission!
 - It ranked first in private main track with ADQA!
 - but its CIDEr score was very low
 - It describes the right events in the wrong words
 - Did not use names often enough
 - Vocabulary/style used was sometimes not typical of ADs

GT: Monday chases Wiley as he pedals the bike lengthwise over the trunks, roofs, and hoods of cars parked in a row.

PRED: A man on a bike jumps between cars in a parking garage and crashes.

CIDEr: 0.23

Presentation of the Main Track Winner's Team

Team Turismo



Thank you!

CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)