

ASSIGNMENT – 04

--Sudha Sree, Yerramsetty

A2)

Models Compared:

- T5 (Text-to-Text Transfer Transformer)
- DistilBART (Distilled BART)

Evaluation Methodology:

Five news articles were selected from a set of generated summaries. Each article was evaluated using both T5 and DistilBART-generated summaries. Four Likert-scale questions (rated 1 to 5) assessed fluency, key point retention, factual consistency, and overall summary quality.

Article 1: Death of Robert H. Schuller: "(CNN)The Rev. Robert H. Schuller, California televangelist and founder of the television ministry "Hour of Power," died Thursday, according to his family. He was 88 years old. Schuller, also the founder of Crystal Cathedral megachurch, had been diagnosed with esophageal cancer in August 2013, a release from "Hour of Power" said. "My father-in-law passed away peacefully early this morning. He was a great Dad and a great man of God," said Schuller's daughter-in-law, Donna Schuller, in a Twitter message. Schuller's life followed an almost Shakespearean arc. He was born in a Iowa farmhouse without running water and longed to preach from his earliest days. In his autobiography, "Prayer: My Soul's Adventure with God," he described standing alone by a river and picturing himself delivering sermons to a rapt congregation. After attending a Hope College and Western Theological Seminary in Michigan, he met his wife of more than 60 years, Arvella, while preaching at her church (she was the organist). With their young family in tow, the Schullers caravanned west to California, where he rented a drive-in theater and preached from the roof of the snack bar. It was beneath the dignity of Christian ministry, some local pastors huffed. The "passion pits" where teenagers necked was no place for the gospel. Schuller was undeterred, and he quickly outgrew the drive-in. He called the explosive growth of his tiny congregation a "miracle," though his many mainstream critics had other names for it. His confident, breezy version of

Christianity -- too breezy, by some estimations -- drew hordes of seekers and lapsed Christians who were put off by the hellfire fulminations of many post-War American preachers. Schuller sold a softer, gentler message, which borrowed heavily, he acknowledged, from the father of the feel-good gospel, Norman Vincent Peale. He preached not to convert or condemn people, but to encourage them, a sentiment he called "possibility thinking." People loved it.

"Evangelicalism at its best wants to be innovative and reach people," said Timothy Larsen, a professor of Christian thought at Wheaton College in Illinois. "And Schuller was a master at that." "What he got right is that the gospel is good news," Larsen continued. "And he preached an uplifting message about personal transformation and uplift and hope." Some of Schuller's favored phrases, though, struck others as cornpone Christianity. "Turn your hurt into a halo?" said Randall Balmer, a professor of American religious history at Dartmouth College, citing one such phrase. "That's pretty weak tea." Still, Balmer gives Schuller some credit. "It may be bad theology, but it's brilliant marketing." In 1970, Schuller began broadcasting "Hour of Power," believed to be one of the first, if not the very first, Sunday service to be shown regularly on television. With his genial smile, priestly robes and gray hair, he looked and talked like a guy who wanted nothing more than to see his flock succeed. The show, which ran for decades, reached millions, making Schuller a televangelist before the term became tarnished by the sins of his many successors. Schuller's crowning achievement, at least architecturally, still stands in Orange County, California, though it is now owned by the Roman Catholic Church. The Crystal Cathedral, a great gleaming edifice with 10,000 glass panels, gave worshipers a look at the clouds that house the heavens, while Schuller preached in the pulpit below. The message was clear to many: The road to the former ran through the latter. During the 1980s and 1990s, Schuller's star continued to rise, with presidents stopping by the Crystal Cathedral -- often during campaigns, it should be said -- and future megachurch pastors like Rick Warren and Bill Hybels seeking his advice. As Schuller aged, though, his family was beset by a succession scandal straight from the pages of "King Lear." He tried to install his only son, Bobby Jr., as pastor of Crystal Cathedral. But the preaching styles of father and son were too different for the congregation -- measured at times at 10,000 strong -- to countenance. Bobby Schuller Jr. left "Hour of Power" and the pulpit at Crystal Cathedral after a short time. As the family searched for a new successor and tussled over finances, viewers and donations to the church and its television show dropped precipitously. Crystal Cathedral Ministries filed for bankruptcy in 2010, citing debts of more than \$43 million, according to The Associated Press. Schuller's empire, which once soared as high as his glassy cathedral, had fallen to dust. Eventually, Schuller's grandson, also named Bobby, took over "Hour of Power,"

though at a different church. In a statement on Thursday, the younger Schuller recalled standing atop Crystal Cathedral's 12-story Tower of Hope with his grandfather as they surveyed the surrounding landscape. "You could see the whole world from there," he said. People we've lost in 2015 . CNN's Stella Chan reported from Los Angeles."

T5 Summary:

"the rev. Robert H. Schuller was diagnosed with esophageal cancer in august 2013 . he was a televangelist and founder of the tv ministry "Hour of Power" his daughter-in-law, Donna Schuller, says he passed away peacefully ."

DistilBART Summary:

" The Rev. Robert H. Schuller was diagnosed with esophageal cancer in August 2013 . He was also the founder of Crystal Cathedral megachurch in Orange County, California ."

Metrics:

1. Fluency – 4 for T5 because it's grammatically correct and well-structured, 5 for DistilBART because it's clear and fluent without any awkward phrasing.
2. Key Points – 3 for T5 because it covers the cancer diagnosis, "Hour of Power," and his peaceful passing, but misses the broader context: his megachurch legacy, theological style, and posthumous issues like the bankruptcy and succession drama, 4 for DistilBART because it adds one more key detail (Crystal Cathedral founder), but still omits much of the rich narrative about Schuller's influence, legacy, and downfall.
3. Factual Consistency - 5 for T5, 5 for DistilBART because both are completely factually correct.
4. Overall Quality - 4 for T5 because it's solid, but the scope is narrow, and it leaves out context that would make the summary more reflective of the article's tone and arc, 4 for DistilBART because it's slightly more informative, thanks to the Crystal Cathedral mention, but still very limited considering the depth and narrative richness of the full article.

Article 2: Michele Bachmann's Obama Comment: "(CNN)Michele Bachmann is comparing President Obama to the co-pilot of the doomed Germanwings flight. "With his Iran deal, Barack Obama is for the 300 million souls of the United States what Andreas Lubitz was for the 150 souls on the German Wings flight - a deranged pilot flying his entire nation into the rocks," the Minnesota Republican and former representative wrote in a Facebook comment posted March 31. "After

the fact, among the smoldering remains of American cities, the shocked survivors will ask, why did he do it?" Andreas Lubitz, the co-pilot of Germanwings Flight 9525, is accused by authorities of deliberately crashing the plane in the French Alps. He died in the crash along with 149 other crew and passengers. The motive of the March 24 crash is under investigation, though investigators are looking in to whether Lubitz feared a medical condition would cause him to lose his pilot's license. Many comments posted on her Facebook page blasted the former representative. Melissa Coca wrote, "Comparing this tragedy to anything is moronic and despicable." Michael J Pristash wrote, "Your allusion is so inappropriate and divisive, not to mention disrespectful on so many levels. Shame on you." Some also accused her of taking desperate measures to stay in the public eye. Lynda Anderson wrote, "Posting outrageous things in a pathetic attempt to stay relevant?" Negotiations are coming down to the wire between Iran, the United States and other nations on restricting Tehran's nuclear program to prevent the ability to develop an atomic bomb. One deadline passed Tuesday, but there is a June 30 deadline for a comprehensive deal -- with all technical and diplomatic impasses fully worked out. Bachmann is no stranger to voicing her opinion on the President's dealing with Iran, personally telling him to "bomb Iran" during the 2014 White House Christmas Party. "I turned to the president and I said, something to the effect of, 'Mr. President, you need to bomb the Iranian nuclear facilities, because if you don't, Iran will have a nuclear weapon on your watch and the course of world history will change,'" she told the Washington Free Beacon. The congresswoman, who sought the GOP presidential nomination in 2012, said Obama had a "condescending smile on his face and laughed at me." She said he told her: "Well Michele, it's just not that easy.""

T5 Summary:

"the motive of the crash is under investigation, though investigators are looking in to whether Lubitz feared a medical condition would cause him to lose his pilot's license .
"your allusion is so inappropriate and divisive, not to mention disrespectful on so many levels. Shame on you," she says ."

DistilBART Summary:

" Michele Bachmann compares President Obama to the co-pilot of the doomed Germanwings flight . Bachmann: "With his Iran deal, Barack Obama is for the 300 million souls of the United States""

Metrics:

1. Fluency – 4 for T5 because it's mostly clear, but the quotes feel somewhat out of context and disjointed, 5 for DistilBART because it's smooth and grammatical, though it ends abruptly.
2. Key Points – 3 for T5 because it omits the central point, 4 for DistilBART because captures the main, controversial claim about Obama and the Iran deal (it doesn't fully elaborate, but it gets to the heart of the article).
3. Factual Consistency - 5 for T5, 5 for DistilBART because both are accurate.
4. Overall Quality - 4 for T5 because the summary is too focused on the Germanwings crash context and neglects Bachmann's political commentary, which is what the article is actually about, 4 for DistilBART since despite being brief, this summary effectively identifies the main controversy, giving it better alignment with the article's purpose.

Article 3: New Research on the Deaths of Anne Frank and Her Sister Margot:
"(CNN)Seventy years ago, Anne Frank died of typhus in a Nazi concentration camp at the age of 15. Just two weeks after her supposed death on March 31, 1945, the Bergen-Belsen concentration camp where she had been imprisoned was liberated -- timing that showed how close the Jewish diarist had been to surviving the Holocaust. But new research released by the Anne Frank House shows that Anne and her older sister, Margot Frank, died at least a month earlier than previously thought. Researchers re-examined archives of the Red Cross, the International Training Service and the Bergen-Belsen Memorial, along with testimonies of survivors. They concluded that Anne and Margot probably did not survive to March 1945 -- contradicting the date of death which had previously been determined by Dutch authorities. In 1944, Anne and seven others hiding in the Amsterdam secret annex were arrested and sent to the Auschwitz-Birkenau concentration camp. Anne Frank's final entry . That same year, Anne and Margot were separated from their mother and sent away to work as slave labor at the Bergen-Belsen camp in Germany. Days at the camp were filled with terror and dread, witnesses said. The sisters stayed in a section of the overcrowded camp with no lighting, little water and no latrine. They slept on lice-ridden straw and violent storms shredded the tents, according to the researchers. Like the other prisoners, the sisters endured long hours at roll call. Her classmate, Nannette Blitz, recalled seeing Anne there in December 1944: "She was no more than a skeleton by then. She was wrapped in a blanket; she couldn't bear to wear her clothes anymore because they were crawling with lice." Listen to Anne Frank's friends describe her concentration camp experience . As the Russians advanced further, the Bergen-Belsen concentration camp became even more crowded, bringing more disease. A deadly typhus outbreak caused thousands to die each

day. Typhus is an infectious disease caused by lice that breaks out in places with poor hygiene. The disease causes high fever, chills and skin eruptions. "Because of the lice infesting the bedstraw and her clothes, Anne was exposed to the main carrier of epidemic typhus for an extended period," museum researchers wrote. They concluded that it's unlikely the sisters survived until March, because witnesses at the camp said the sisters both had symptoms before February 7. "Most deaths caused by typhus occur around twelve days after the first symptoms appear," wrote authors Erika Prins and Gertjan Broek. The exact dates of death for Anne and Margot remain unclear. Margot died before Anne. "Anne never gave up hope," said Blitz, her friend. "She was absolutely convinced she would survive." Her diary endures as one of the world's most popular books. Read more about Anne Frank's cousin, a keeper of her legacy ."

T5 Summary:

"typhus is an infectious disease caused by lice that breaks out in places with poor hygiene . the disease causes high fever, chills and skin eruptions . researchers re-examine archives of the red cross, the international training service and the Bergen-Belsen Memorial ."

DistilBART Summary:

" Researchers re-examined archives of the Red Cross, the International Training Service and the Bergen-Belsen Memorial . They concluded that Anne and Margot Frank probably did not survive to March 1945 . Days at the camp were filled with terror and dread, witnesses said ."

Metrics:

1. Fluency – 5 for T5, 5 for DistilBART because both are well-written, grammatically sound and clear.
 2. Key Points – 4 for T5 because it covers the research and disease context but doesn't delve into the details of the sisters' experience in the camp as well as DistilBART, 5 for DistilBART because it provides a better overview of the research findings and the conditions the sisters endured.
 3. Factual Consistency - 5 for T5, 5 for DistilBART because both are factually accurate and do not hallucinate.
 4. Overall Quality - 4 for T5 because it's focused more on the scientific details of the disease, missing some emotional context, 5 for DistilBART because it has a stronger coverage of both the scientific findings and the emotional backdrop of the sisters' suffering.
-

Article 4: Bob Barker Returns to *The Price Is Right*: "(CNN)For the first time in eight years, a TV legend returned to doing what he does best. Contestants told to "come on down!" on the April 1 edition of "The Price Is Right" encountered not host Drew Carey but another familiar face in charge of the proceedings. Instead, there was Bob Barker, who hosted the TV game show for 35 years before stepping down in 2007. Looking spry at 91, Barker handled the first price-guessing game of the show, the classic "Lucky Seven," before turning hosting duties over to Carey, who finished up. Despite being away from the show for most of the past eight years, Barker didn't seem to miss a beat."

T5 Summary:

"contestants told to "come on down!" on the April 1 edition of "The Price Is Right" they encountered not host Drew Carey but another familiar face in charge of the proceedings . Bob Barker hosted the show for 35 years before stepping down in 2007 ."

DistilBART Summary:

" Bob Barker returned to "The Price Is Right" for the first time in eight years . The 91-year-old hosted the show for 35 years before stepping down in 2007 ."

Metrics:

1. Fluency – 4 for T5 because it's almost perfect, but the first sentence is a little long and clunky/confusing, 5 for DistilBART because it's fluent, clean and easy to read.
2. Key Points – 4 for T5 because it mentions the surprise appearance, the date (April 1), and Drew Carey's absence, but doesn't explicitly say Bob only hosted one game, 4 for DistilBART because it covers the return and historical context well, but it misses details like the segment he hosted and that Carey finished the show.
3. Factual Consistency - 5 for T5, 5 for DistilBART because all facts align with article for both summaries.
4. Overall Quality - 4 for T5 because it's slightly more detailed and vivid, 5 for DistilBART because it's more concise and elegant, though a bit less informative overall.

Article 5: Ebola Crisis in West Africa: "(CNN)Five Americans who were monitored for three weeks at an Omaha, Nebraska, hospital after being exposed to Ebola in West Africa have been released, a Nebraska Medicine spokesman said in an email Wednesday. One of the five had a heart-related issue on Saturday and has been discharged but hasn't left the area, Taylor Wilson wrote. The others have already

gone home. They were exposed to Ebola in Sierra Leone in March, but none developed the deadly virus. They are clinicians for Partners in Health, a Boston-based aid group. They all had contact with a colleague who was diagnosed with the disease and is being treated at the National Institutes of Health in Bethesda, Maryland. As of Monday, that health care worker is in fair condition. The Centers for Disease Control and Prevention in Atlanta has said the last of 17 patients who were being monitored are expected to be released by Thursday. More than 10,000 people have died in a West African epidemic of Ebola that dates to December 2013, according to the World Health Organization. Almost all the deaths have been in Guinea, Liberia and Sierra Leone. Ebola is spread by direct contact with the bodily fluids of an infected person.", T-5 summary = "five americans who were monitored for three weeks have been released, a spokesman says . one of the five had a heart-related issue on Saturday and has been discharged . none of them developed the deadly virus ."

T5 Summary:

"five americans who were monitored for three weeks have been released, a spokesman says . one of the five had a heart-related issue on Saturday and has been discharged . none of them developed the deadly virus ."

DistilBART Summary:

" The five were exposed to Ebola in Sierra Leone in March, but none developed the deadly virus . One of the five had a heart-related issue on Saturday and has been discharged but hasn't left the area ."

Metrics:

1. Fluency – 4 for T5 because it's fluent and grammatical, though the sentence structure is slightly repetitive, 5 for DistilBART because it reads smoothly and is well-constructed.
2. Key Points – 4 for T5 because it captures the core facts, but misses organizational affiliation and broader context, 4 for DistilBART because it also includes the key facts about exposure, monitoring, and the heart-related issue, but like T5, it omits mention of Partners in Health and the broader CDC or WHO context.
3. Factual Consistency - 5 for T5, 5 for DistilBART since both are faithful to the original article.
4. Overall Quality - 4 for T5 because it's a clean summary of key facts, though a bit dry and missing context, 5 for DistilBART because it's slightly more readable and polished, with good compression of info into just two sentences.

Reflection Question:

Which Evaluation Method Is Better: ROUGE-1 or Manual Evaluation?

Answer:

While ROUGE-1 provides a fast, quantitative comparison of n-gram overlap between generated and reference summaries, manual evaluation offers a deeper and more nuanced understanding of summary quality. This is particularly important for:

- **Capturing paraphrases** and non-exact matches
- **Recognizing factual correctness** even when phrasing differs
- **Evaluating stylistic fluency and readability**, which ROUGE does not account for

In this evaluation, several T5 summaries had decent ROUGE-1 scores but missed key context or framing, while DistilBART summaries, although not always longer, often scored better manually due to better focus, tone, and accuracy. Therefore, manual evaluation is a more reliable indicator of summary quality, especially when summaries are used in sensitive or public-facing contexts. ROUGE is useful for benchmarking at scale, but human review is critical for final assessments.