# Finding Mistakes in Basketball Stories: Text 10

Thanks very much for helping with our project! Just to give some background, lots of people are excited by the potential of “deep learning” techniques in artificial intelligence, which have been used in self-driving cars and Google Translate, for example. However deep learning has not been very successful in our research area of “Natural Language Generation”, which is about getting computers to write stories from data.

We are trying to develop better ways of using deep learning in computer systems that write stories from data, and as part of this project we’re trying to figure out how to best detect accuracy problems in computer-generated stories. I.e., cases where the stories say things that aren’t true. We need to find these problems before we can fix them.

As part of this work, we have developed some guidelines for marking up mistakes in computer-generated summaries of basketball games, based on box scores from these games. Focusing on cases where the reports are wrong, not grammar or spelling errors.

What we’d like you to do is to mark-up mistakes in a computer-generated basketball summary, using our guidelines. We’re interested in the mistakes that you find, and also in whether the guidelines make sense. If you find that the guidelines are confusing or don’t work, please let me know!

We’d also like to know some basic information about you

Following this intro, you’ll see

* our guidelines on pages 2 and 3
* an example basketball story which we have marked up, on pages 4 and 5
* the story we would like you to mark up, on page 6. We give links to basketball-reference.com for the box score information as well as the season schedule for each team (such that you can find other games which might be mentioned in the text).
* Space for any additional comments you may have and page 7.

Please mark up the game summary on page 6 in a similar way to the example on pages 4 and send us the document.

Thanks for your help!

Ehud Reiter

## Mark-up guidelines

We’ve given you some basketball game stories produced by “deep learning” AI systems, and also links to box score information on basketball-reference.com about the games (the stories focus on box scores, they don’t talk about individual goals, penalties, etc). We’ve also given links to each teams’ season schedule for season information (some of the stories say where the next game will be). Please read through the stories and mark up cases where

* numbers are wrong
* names (players, teams, cities, etc) are wrong
* words are wrong
* context means people will misunderstand a sentence
* facts which are not checkable
* other cases where the story says something which is not true

We give more information below about these types of errors.

Please mark up the wrong numbers, names, etc by putting them in red. If you’re colour-blind, you can underline them instead. Also please add a note below the story for each mistake; the note should explain the mistake and say which type it is. There is an example on pages 4 and 5.

### Number mistakes

Numbers mistakes are incorrect numbers. For example

* “10-point victory” when margin of victory was 11 pts
* “six players reached double figures” when only four players did so.

Please mark-up the wrong number by putting it in red or underlining it. It doesn’t matter whether the number is digits (such as 10) or written as a word (such as six).

### Name mistakes

Name mistakes are errors in things that have names. This includes people, cities, teams, stadiums, and days of week. If a word (other than “I”) is always capitalised, it is probably a name. For example

* “on Monday” when game was played on Wednesday
* “Talking Stick Resort Arena” when game was played in US Airways Arena.

Please mark-up the wrong name by putting it in red or underlining it.

### Word mistakes

Word mistakes are incorrect or inappropriate words which are not names or numbers. For example

* “out-scored the Suns” when the Suns had a higher score in this period
* “off the bench” for a player who was on the starting team
* “strong first half” when team did poorly in first half

Please mark-up the wrong word by putting it in red or underlining it.

We treat mistakes in fixed phrases such as “off the bench” as word mistakes (the AI systems treat fixed phrases in the same way as they treat words).

The appendix defines some words and phrases which appear in descriptions of basketball games. It is incomplete, let me know if you have any suggestions or corrections for it.

### Context mistakes

Context mistakes occur when people reading a sentence are likely to misinterpret it because of its context, even if the sentence is literally true. For example

* “The Suns had six players reach double figures in points. Mike Conley led the way with 24 points.” This is a context error because Conley played for the other team (not the Suns). Ie, Conley did score 24 points, but the context implies he played for the Suns, which is wrong.

For the mark-up, try to find the thing which will be misinterpreted (as above), and put it in red or underline it.

### Facts which are not checkable

Some facts will not be practical to check. We do not expect you to look forward or back further than 4 games to check a statement. For example:

* Please do check: [not in example game] “It was his second double-double in a row, …”. You can find the players per-game history by clicking on their name in the box score, then selecting the appropriate season. Please also check each teams’ next game details.
* Do not check [not in example game] “The Wizards came into this game as the worst rebounding team in the NBA.”.

For the mark-up, highlight the text which cannot be checked, and put it in red or underline it. Also indicate that it is the “not checkable” category.

### Other mistakes

This includes errors which don’t fit into one of the above categories, and also cases where the story has so many mistakes that it is difficult to individually mark them. For example

* [not in example game] “to give the Suns a commanding Game 1 loss to give the Suns a 118 - 0 record in the Eastern\_Conference 's first playoff series with at least the Eastern\_Conference win in Game 5” when the game is not a playoff game, Suns are not in Eastern Conference, 118-0 is impossible as a team record, etc

Try to mark-up the wrong text by putting it in red or underlining it. We can’t give precise instructions because the “other” category is very broad.

### Notes

If a sentence is repeated, this is not an accuracy mistake, unless of course the sentence contains errors. For example

* “The Sun’s next game will be on Friday, while the Sun’s next game will be on Friday” is a bizarre thing to say, but it is not an accuracy mistake (assuming that the next game will in fact be on Friday).

If there are multiple ways in which you can annotate a sentence for errors, choose the one with the fewest total errors. For Example, when choosing between [not in example game]:

* “Lou Williams led the team in scoring, dropping 30 points, six rebounds and seven assists.”.
* “Lou Williams led the team in scoring, dropping 30 points, six rebounds and seven assists.”.

The first annotation is preferred as the text describes the exact statistics of another player, Solomon Hill.

## Example marked-up story

Game played on 5-Nov-2014, between Memphis Grizzlies and Phoenix Suns.

Main box score data

* <https://www.basketball-reference.com/boxscores/201411050PHO.html>

Other useful data

* Home team season schedule: <https://www.basketball-reference.com/teams/PHO/2015_games.html>
* Visiting team season schedule: <https://www.basketball-reference.com/teams/MEM/2015_games.html>
* Online calendar: <https://www.timeanddate.com/calendar/monthly.html?year=2014&month=11&country=1>

Story, with errors in underlined red

The Memphis Grizzlies ( 5 - 2 ) defeated the Phoenix Suns ( 3 - 2 ) Monday 1 - 2 at the Talking Stick Resort Arena in Phoenix . The Grizzlies had a strong first half where they out - scored the Suns 59 - 42 , to coast to a 10 - point victory in front of their home crowd . The Grizzlies were led by Isaiah Thomas , who scored 15 points ( 4 - 10 FG , 1 - 4 3Pt , 6 - 6 FT ) . He also had six rebounds and five assists in 26 minutes . Eric Bledsoe had 23 points ( 9 - 12 FG , 2 - 2 3Pt , 3 - 4 FT ) , five rebounds and four assists , while Bledsoe added 24 points ( 9 - 14 FG , 2 - 4 3Pt , 3 - 4 FT ) , five rebounds and four assists . The Suns had six players reach double figures in points . Mike Conley led the way with 24 points ( 9 - 14 FG , 3 - 4 3Pt , 3 - 5 FT ) and 11 assists , while Tony Allen chipped in with nine points ( 4 - 6 FG , 1 - 1 FT ) and a pair of assists . The Suns had six players reach double figures in points in this one. Tony Allen had nine points ( 4 - 6 FG , 1 - 1 FT ) and a pair of assists off the bench . The Suns ' next game will be on the road against the Boston Celtics on Friday , while the Suns will be at home against the Portland Trail Blazers on Friday .

LIST OF MISTAKES

5 - 2 - should be 5-0 [number]

Monday – Wednesday [name]

1 - 2 - score was 102-91 [number]

Talking Stick Resort Arena - that is the name now, but in 2014 the Stadium was called US Airways Arena [name]

strong – first half was not strong [word]

outscored – Suns outscored Grizzlies [word]

59-42 – actual score was 46-52 [number]

59-42 – actual score was 46-52 [number]

coast – they had to catch up from behind [word]

10 point – should be 11 point [number]

home – Game was in Phoenix [word]

led – Thomas did not lead the Grizzlies [word]

Isaiah Thomas – Thomas played for Suns, not Grizzlies [context]

six – one rebound [number]

five – two assists [number]

Eric Bledsoe – Bledsoe played for the Suns, not Grizzlies [context]

24, 14, 4 – correct figures for Bledsoe are in previous clause (23, 12, 2) [number]

24, 14, 4 – correct figures for Bledsoe are in previous clause (23, 12, 2) [number]

24, 14, 4 – correct figures for Bledsoe are in previous clause (23, 12, 2) [number]

six – only four Sun players reached double figures [number]

Mike Conley – Conley plays for the Grizzlies, not Suns [context]

Tony Allen – Allen plays for the Grizzlies, not Suns. [context]

Pair – one assist [number]

six – only four Grizzly players reached double figures [number]

off the bench – Allen was a starter [word]

on the road – home game [word]

Boston Celtics – next game is against Sacramento [name]

Portland Trail Blazers – next game is against Sacramento [name]

## Story to be marked-up

Please mark up the following story, links to the statistics for the game on basketball-reference.com, as well as each teams’ season schedule are included below to assist you.

Main box score data

* <https://www.basketball-reference.com/boxscores/201701040ORL.html>

Other useful data

* Home team season schedule: <https://www.basketball-reference.com/teams/ORL/2017_games.html>
* Visiting team season schedule: <https://www.basketball-reference.com/teams/ATL/2017_games.html>
* Online calendar: <https://www.timeanddate.com/calendar/monthly.html?year=2017&month=1&country=1>

STORY

The Atlanta Hawks defeated the host Orlando Magic, 111-92, at Amway Center on Friday1. In a battle between two struggling teams2, the Hawks were able to prevail with a huge road win. In fact, there were 11 ties3 and 11 lead changes4 in total, with neither team leading by more than 11 points5. Defense was key for Atlanta, as they held Orlando to 43 percent from the field and 53 percent from three-point range6. Rebounding was a huge factor7 as well, with the Hawks recording 27 assists to 12 turnovers, while the Magic committed 17 turnovers to just 19 assists. The Hawks (19-16) have now lost four of their last five games, as they continue to struggle8. Dennis Schroder was the player of the game, as he tallied 18 points, six rebounds and seven assists on 7-of-13 shooting. Dwight Howard recorded a double-double, totaling 13 points and 12 rebounds. Kent Bazemore was the only other starter in double figures9, as he totaled 17 points and four assists. On deck for Atlanta is a home matchup10 against the New Orleans Pelicans on Saturday11. The Magic (16-21) have been struggling mightily recently, but this was a huge win12. Nikola Vucevic was the player of the game13, as he collected 10 points, 12 rebounds14 and six assists. Evan Fournier was second on the team15, as he accrued 14 points, one rebound and one assist. Nikola Vucevic recorded a double-double16, totaling 10 points and 10 rebounds17. Evan Fournier was the only other Magic player in double figures18, as he amassed 14 points and one rebound. Orlando will look to keep rolling19 on Friday in a home matchup against the Toronto Raptors20.

LIST OF MISTAKES

1. Friday – The game took place on a Wednesday. [NAME]
2. Two struggling teams – Atlanta went into the game with an overall winning record (18-16) and were on a 3-game win streak. That is not struggling, in my opinion. [WORD]
3. 11 ties – There were only three ties. [NUMBER]
4. 11 lead changes – There were only two lead changes. [NUMBER]
5. neither team leading by more than 11 points – The final score shows a 19-point difference, so this is clearly wrong. [NUMBER]
6. 53 percent from three-point range – Orlando shot 32% from 3-point range. 53% is the 3P% stat for Atlanta, not Orlando. [NUMBER]
7. Rebounding was a huge factor – Orlando were better than Atlanta on rebounds (40-39) but still lost by 19 points, so rebounding was not a huge factor. Also, the sentence goes on to give stats that, though correct, have nothing to do with rebounding. Assist and turnover stats are good for evaluating how effectively a team uses possession of the ball. [WORD]
8. lost four of their last five games, as they continue to struggle– Following this win, Atlanta had won four of their last five games. “Lost” is obviously the wrong word, but there is also a wider evaluative error (“continue to struggle”) relating to the team’s form. [OTHER]
9. Kent Bazemore was the only other starter in double figures– Assuming we’re talking about scoring here (which is my inference), Paul Millsap also started for Atlanta and reached double figures (16 points). Dwight Howard and Dennis Schroder achieved the same feat; but, since they are mentioned earlier in the report, Millsap is the only omission. [OTHER]
10. home matchup– Atlanta’s next game is at New Orleans, not at home. [WORD]
11. Saturday – The next game is on a Thursday (Jan 5, 2017). [NAME]
12. struggling mightily recently, but this was a huge win– The main problem here is that Orlando didn’t win; however, the sentence is also inaccurate because it suggests the Magic were in extremely poor form going into the game when in fact their form (3 wins in last 5 games) wasn’t bad. [OTHER]
13. Nikola Vucevic was the player of the game– The report previously described Dennis Schroder as the player in the game, so the report contains two mutually exclusive statements. Even if you gave the sentence greater specificity (e.g. “Nikola Vucevic was Orlando’s player of the game”) it would still be dubious as several of Vucevic’s teammates have better stats. [OTHER]
14. 12 rebounds – Vucevic got 6 rebounds. [NUMBER]
15. Evan Fournier was second on the team – Second in what respect? He did well in scoring, but I’d put him joint third in that category, with his teammates Payton and Gordon joint first with 15 points each. Neither Payton nor Gordon are mentioned, which makes me think this part of the report is based on an analysis that omitted the Orlando starters’ contributions altogether. [OTHER]
16. Nikola Vucevic recorded a double-double – Vucevic didn’t record a double-double. He achieved double digits in one offensive category only. [OTHER]
17. 10 rebounds – Vucevic got 6 rebounds. [NUMBER]
18. Evan Fournier was the only other Magic player in double figures – Assuming this is a reference to points scored, four other Magic players achieved double figures: Payton, Gordon, Vucevic, and Green. Vucevic’s double-digit scoring is mentioned earlier in the report, but Payton, Gordon, and Green are omissions. [OTHER]
19. keep rolling – This phrase implies that Orlando won. They lost. Badly. [WORD]
20. Toronto Raptors– Should be Houston Rockets. [NAME]

## Comments

If you have any comments about this exercise, please write them below.